

**UNITED STATES PATENT AND TRADEMARK OFFICE**

---

**BEFORE THE PATENT TRIAL AND APPEAL BOARD**

---

TIKTOK INC.,

Petitioner,

v.

SHOPSEE, INC.,

Patent Owner.

---

U.S. Patent No. 11,134,316

“Integrated Shopping Within Long-Form Entertainment”

---

DECLARATION OF KEVIN C. ALMEROOTH, PH.D., IN SUPPORT OF  
PETITION FOR *INTER PARTES* REVIEW OF  
U.S. PATENT NO. 11,134,316

**TABLE OF CONTENTS**

|      |                                                       |    |
|------|-------------------------------------------------------|----|
| I.   | Introduction.....                                     | 1  |
| A.   | Qualifications .....                                  | 1  |
| 1.   | Educational Background.....                           | 1  |
| 2.   | Career .....                                          | 2  |
| 3.   | Other Relevant Qualifications.....                    | 9  |
| B.   | Previous Expert Witness Testimony .....               | 10 |
| C.   | Preparation for this Declaration .....                | 10 |
| II.  | Legal Understanding.....                              | 11 |
| A.   | Claim Construction.....                               | 12 |
| B.   | Anticipation .....                                    | 12 |
| C.   | Obviousness.....                                      | 13 |
| III. | The '316 Patent.....                                  | 14 |
| A.   | Priority Date .....                                   | 14 |
| B.   | Specification.....                                    | 14 |
| C.   | Prosecution History of the '316 Patent (EX1004) ..... | 16 |
| D.   | Level of Ordinary Skill in the Art.....               | 18 |
| E.   | Claim Construction.....                               | 19 |
| F.   | Challenged Claims .....                               | 23 |
| IV.  | State of the Art.....                                 | 30 |

|    |                                                                    |     |
|----|--------------------------------------------------------------------|-----|
| V. | Grounds.....                                                       | 32  |
| A. | <u>Ground 1: Ohliger</u> .....                                     | 32  |
| 1. | Background .....                                                   | 32  |
| 2. | Analysis.....                                                      | 35  |
| B. | Ground 2: Ohliger and Sakhartov .....                              | 166 |
| 1. | Background .....                                                   | 166 |
| 2. | Proposed Combination.....                                          | 166 |
| 3. | Motivation to Combine & Reasonable Expectation of Success<br>..... | 166 |
| 4. | Analysis.....                                                      | 174 |
| C. | Ground 3: Ohliger and Kaiserlian .....                             | 176 |
| 1. | Background .....                                                   | 176 |
| 2. | Proposed Combination.....                                          | 176 |
| 3. | Motivation to Combine & Reasonable Expectation of Success<br>..... | 177 |
| 4. | Analysis.....                                                      | 185 |
| D. | Ground 4: Ohliger and Tapissier.....                               | 190 |
| 1. | Background .....                                                   | 190 |
| 2. | Proposed Combination.....                                          | 190 |
| 3. | Motivation to Combine & Reasonable Expectation of Success<br>..... | 191 |
| 4. | Analysis.....                                                      | 199 |
| E. | Ground 5: Ohliger, Sakhartov, and Kaiserlian or Tapissier.....     | 205 |

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

VI. Secondary Considerations of Non-Obviousness .....206

VII. Conclusion .....206

**LIST OF EXHIBITS**

| <b>Exhibit</b> | <b>Description</b>                                                                                                                            |
|----------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| EX1001         | U.S. Patent No. 11,134,316 (“’316 patent”)                                                                                                    |
| EX1002         | Declaration of Kevin C. Almeroth.                                                                                                             |
| EX1003         | Curriculum Vitae of Kevin C. Almeroth                                                                                                         |
| EX1004         | Prosecution History for U.S. Pat. No. 11,134,316                                                                                              |
| EX1005         | Patent Owner’s District Court Infringement Chart for ’316 Patent                                                                              |
| EX1006         | U.S. Pat. No. 11,432,046 (“Ohliger”)                                                                                                          |
| EX1007         | E.P. 1,425,918 (“Tapissier”)                                                                                                                  |
| EX1008         | Constance Steinkuehler, “Massively Multiplayer Online Gaming as a Constellation of Literacy Practices,” 4 E-Learning 3 (2007)                 |
| EX1009         | U.S. Pat. Appl. Pub. 2016/0232579 (“Fahnestock”)                                                                                              |
| EX1010         | U.S. Pat. No. 9,277,164 (“Lane”)                                                                                                              |
| EX1011         | U.S. Pat. No. 11,432,046 – Drawings – Other than Black and White Line Drawings, Fig. 1-12 (June 13, 2016)                                     |
| EX1012         | U.S. Pat. No. 11,432,046 – Drawings – Other than Black and White Line Drawings, Fig. 13-29 (June 13, 2016)                                    |
| EX1013         | U.S. Pat. No. 8,166,500 (“Sakhartov “)                                                                                                        |
| EX1014         | U.S. Pat. No. 8,332,477 (“Kaiserlian”)                                                                                                        |
| EX1015         | SCORE Placeholder Sheet for IFW Content, Prosecution History for U.S. Pat. No. 11,432,046 (June 13, 2016)                                     |
| EX1016         | TikTok Opening Claim Construction Brief (Dkt. 66)                                                                                             |
| EX1017         | ShopSee Claim Construction Response Brief (Dkt. 73)                                                                                           |
| EX1018         | Michael Isard et al., <i>CONDENSATION—Conditional Density Propagation for Visual Tracking</i> , 29 Int’l J. of Comput. Vision 1, 5-28 (1998). |
| EX1019         | U.S. Patent No. 11,368,514 (“Boffa”)                                                                                                          |
| EX1020         | U.S. Pat. Appl. Pub. 2008/0294694 (“Maghfourian”)                                                                                             |

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

| Exhibit | Description                                   |
|---------|-----------------------------------------------|
| EX1021  | U.S. Pat. Appl. Pub 2015/0172563 (“Jakobson”) |

I, Dr. Kevin C. Almeroth, declare as follows:

**I. Introduction**

1. I have been retained by Quinn Emanuel Urquhart & Sullivan, LLP on behalf of the Petitioner TikTok Inc. (“Petitioner”) as an independent expert in this *inter partes* review (this “Proceeding”) before the Patent Trial and Appeal Board of the United States Patent and Trademark Office (the “Board”) to review claims 1-19 (“the challenged claims”) of U.S. Patent No. 11,134,316 (“the ’316 patent”). I have been asked by the Petitioner to assist in evaluating the claims and the disclosure of the ’316 patent.

**A. Qualifications**

2. EX1003 is a true and correct copy of my current CV, which describes my education, patents and publications, employment and research history, and professional activities and awards.

**1. Educational Background**

3. I hold three degrees from the Georgia Institute of Technology: (1) a Bachelor of Science degree in Information and Computer Science (with minors in Economics, Technical Communication, and American Literature) earned in June 1992; (2) a Master of Science degree in Computer Science (with specialization in Networking and Systems) earned in June 1994; and (3) a Doctor of Philosophy

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

(Ph.D.) degree in Computer Science (Dissertation Title: Networking and System Support for the Efficient, Scalable Delivery of Services in Interactive Multimedia System, minor in Telecommunications Public Policy) earned in June 1997. I have taken a wide variety of courses as demonstrated by my minor. My undergraduate degree also included a number of courses more typical of a degree in electrical engineering, including digital logic, signal processing, and telecommunications theory.

## **2. Career**

4. I am a Professor Emeritus in the Department of Computer Science at the University of California, Santa Barbara (UCSB). While active at UCSB, I held faculty appointments and was a founding member of the Computer Engineering (CE) Program, Media Arts and Technology (MAT) Program, and the Technology Management Program (TMP). I was the Associate Director of the Center for Information Technology and Society (CITS) from 1999 to 2012. I have been a faculty member at UCSB since July 1997.

5. One of the major concentrations of my research has been the delivery of multimedia content and data between computing devices, including various network architectures. In my research, I have studied large-scale content delivery systems, and the use of servers located in a variety of geographic locations to provide scalable delivery to hundreds or thousands of users simultaneously. I have also

studied smaller-scale content delivery systems in which content is exchanged between individual computers and portable devices. My work has emphasized the exchange of content more efficiently across computer networks, including the scalable delivery of content to many users, mobile computing, satellite networking, delivering content to mobile devices, and network support for data delivery in wireless networks.

6. In 1992, the initial focus of my research was on the provision of interactive functions (*e.g.*, VCR-style functions like pause, rewind, and fast-forward) for near video-on-demand systems in cable systems; in particular, how to aggregate requests for movies at a cable head-end and then how to satisfy a multitude of requests using one audio/video stream broadcast to multiple receivers simultaneously. This research has continually evolved and resulted in the development of techniques to scalably deliver on-demand content, including audio, video, web documents, and other types of data, through the Internet and over other types of networks, including over cable systems, broadband telephone lines, and satellite links.

7. An important component of my research has been investigating the challenges of communicating multimedia content, including video, between computers and across networks including the Internet. I have worked on a variety of research problems and used a number of systems that were developed to deliver

multimedia content to users. One content-delivery method I have researched is the one-to-many communication facility called “multicast,” first deployed as the Multicast Backbone, a virtual overlay network supporting one-to-many communication. Multicast is one technique that can be used on the Internet to provide streaming media support for complex applications like video-on-demand, distance learning, distributed collaboration, distributed games, and large-scale wireless communication. The delivery of media through multicast often involves using Internet infrastructure, devices and protocols, including protocols for routing and TCP/IP.

8. Starting in 1997, I worked on a project to integrate the streaming media capabilities of the Internet together with the interactivity of the web. I developed a project called the Interactive Multimedia Jukebox (IMJ). Users would visit a web page and select content to view. The content would then be scheduled on one of a number of channels, including delivery to students in Georgia Tech dorms delivered via the campus cable plant. The content of each channel was delivered using multicast communication.

9. More recently, I have also studied issues concerning how users choose content, especially when considering the price of that content. My research has examined how dynamic content pricing can be used to control system load. By raising prices when systems start to become overloaded (*i.e.*, when all available

resources are fully utilized) and reducing prices when system capacity is readily available, users' capacity to pay as well as their willingness can be used as factors in stabilizing the response time of a system. This capability is particularly useful in systems where content is downloaded or streamed on-demand to users.

10. Protecting networks, including their operation and content, has been an underlying theme of my research almost since the beginning of my research career. Starting in 2000, I have been involved in several projects that specifically address security, network protection, and firewalls. After significant background work, a team on which I was a member successfully submitted a \$4.3M grant proposal to the Army Research Office (ARO) at the Department of Defense to propose and develop a high-speed intrusion detection system. Key aspects of the system included associating streams of packets and analyzing them for viruses and other malware. Once the grant was awarded, we spent several years developing and meeting the milestones of the project. A number of my students worked on related projects and published papers on topics ranging from intrusion detection to developing advanced techniques to be incorporated into firewalls. I have also used firewalls, including their associated malware detection features, in developing techniques for the classroom to ensure that students are not distracted by online content.

11. Recent work ties some of the various threads of my past research together. I have investigated content delivery in online social networks and proposed

reputation management systems in large-scale social networks and marketplaces. On the content delivery side, I have looked at issues of caching and cache placement, especially when content being shared and the cache has geographical relevance. We were able to show that effective caching strategies can greatly improve performance and reduce deployment costs. Our work on reputation systems showed that reputations have economic value, and as such, creates a motivation to manipulate reputations. In response, we developed a variety of solutions to protect the integrity of reputations in online social networks. The techniques we developed for content delivery and reputation management were particularly relevant in peer-to-peer communication.

12. My involvement in the research community extends to leadership positions for several academic journals and conferences. I am the co-chair of the Steering Committee for the ACM Network and System Support for Digital Audio and Video (NOSSDAV) workshop and on the Steering Committees for the International Conference on Network Protocols (ICNP), ACM Sigcomm Workshop on Challenged Networks (CHANTS), and IEEE Global Internet (GI) Symposium. I have served or am serving on the Editorial Boards of IEEE/ACM Transactions on Networking, IEEE Transactions on Mobile Computing, IEEE Network, ACM Computers in Entertainment, AACE Journal of Interactive Learning Research (JILR), and ACM Computer Communications Review. I have co-chaired a number

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

of conferences and workshops including the IEEE International Conference on Network Protocols (ICNP), IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), International Conference on Communication Systems and Networks (COMSNETS), IFIP/IEEE International Conference on Management of Multimedia Networks and Services (MMNS), the International Workshop On Wireless Network Measurement (WinMee), ACM Sigcomm Workshop on Challenged Networks (CHANTS), the Network Group Communication (NGC) workshop, and the Global Internet Symposium, and I have served on the program committees for numerous conferences.

13. Furthermore, in the courses I taught at UCSB, a significant portion of my curriculum covered aspects of the Internet and network communication including the physical and data link layers of the Open System Interconnect (OSI) protocol stack, and standardized protocols for communicating across a variety of physical media such as cable systems, telephone lines, wireless, and high-speed Local Area Networks (LANs). The courses I have taught also cover most major topics in Internet communication, including data communication, multimedia encoding, and mobile application design. My research and courses have covered a range of physical infrastructures for delivering content over networks, including cable, Integrated Services Digital Network (ISDN), Ethernet, Asynchronous

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

Transfer Mode (ATM), fiber, and Digital Subscriber Line (DSL). For a complete list of courses I have taught, see my curriculum vitae (EX1004).

14. I co-founded a technology company called Santa Barbara Labs that was working under a sub-contract from the U.S. Air Force to develop very accurate emulation systems for the military's next generation internetwork. Santa Barbara Labs' focus was in developing an emulation platform to test the performance characteristics of the network architecture in the variety of environments in which it was expected to operate, and, in particular, for network services including IPv6, multicast, Quality of Service (QoS), satellite-based communication, and security. Applications for this emulation program included communication of a variety of multimedia-based services, including video conferencing and video-on-demand.

15. In addition to having co-founded a technology company myself, I have worked for, consulted with, and collaborated with companies for nearly 30 years. These companies range from well-established companies to start-ups and include IBM, Hitachi Telecom, Turner Broadcasting System (TBS), Bell South, Digital Fountain, RealNetworks, Intel Research, Cisco Systems, and Lockheed Martin.

16. Additional details about my employment history, fields of expertise, and publications are further included in my CV (EX1003).

### **3. Other Relevant Qualifications**

17. I am a Member of the Association of Computing Machinery (ACM) and a Fellow of the Institute of Electrical and Electronics Engineers (IEEE).

18. As an important component of my research program, I have been involved in the development of academic research into available technology in the market place. One aspect of this work is my involvement in the Internet Engineering Task Force (IETF). The IETF is a large and open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. I have been involved in various IETF groups including many content delivery-related working groups like the Audio Video Transport (AVT) group, the MBone Deployment (MBONED) group, Source Specific Multicast (SSM) group, the Inter-Domain Multicast Routing (IDMR) group, the Reliable Multicast Transport (RMT) group, the Protocol Independent Multicast (PIM) group, etc. I have also served as a member of the Multicast Directorate (MADDOGS), which oversaw the standardization of all things related to multicast in the IETF. Finally, I was the Chair of the Internet2 Multicast Working Group for seven years.

19. I am an author or co-author of approximately 200 technical papers, published software systems, IETF Internet Drafts and IETF Request for Comments (RFCs). A complete list of my publications is in my CV (EX1004).

20. I have been awarded numerous teaching awards, including Computer Science Outstanding Faculty Member (1997-98, 1998-99, 1999-2000, 2004-06, UCSB Spotlight on Excellence Award (2000-01), and UCSB Academic Senate Distinguished Teaching Award (2006-07).

**B. Previous Expert Witness Testimony**

21. The list of recent matters in which I have testified can be found in EX1003.

**C. Preparation for this Declaration**

22. In forming my opinions, I have considered the '316 patent specification, including the Abstract, the figures, and the claim language itself, as would have been understood by a person of ordinary skill in the art as of the priority date of the '316 patent (a "POSITA"). My understanding of "POSITA" and "priority date" are set forth below. I have also reviewed the file history of the '316 patent, the Exhibits that are listed in the list of Exhibits, and any other material cited in this declaration.

23. In forming my opinions, I have relied on my personal knowledge and professional experience, and on the documents and information referenced in this declaration.

24. This declaration explains, based on facts and information available to me to date, the subject matter and opinions related to this Proceeding. As such, I am

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

prepared to provide expert testimony regarding opinions formed resulting from my analysis of the issues considered in this declaration if asked about those issues by the Board or by the private parties' attorneys.

25. Additionally, I may discuss my own work, teachings, and knowledge of the state of the art in the relevant time period. I may rely on handbooks, textbooks, technical literature, and the like to demonstrate the state of the art in the relevant period and the evolution of relevant technologies.

26. Throughout this declaration, I refer to specific pages of the '316 patent and other documents. The citations are intended to be exemplary and are not intended to convey that the citations are the only source of evidence to support the propositions for which they are cited.

27. I am being compensated for my time spent on this matter at a rate of \$850 per hour, and my compensation is in no way contingent upon the outcome of this matter or on the opinions I offer. All of the opinions expressed in this declaration are my own.

## **II. Legal Understanding**

28. In this section, I describe my understanding of certain legal standards that I have relied upon in forming my opinions set forth in this declaration. I have been informed of these legal standards by Petitioner's attorneys. I am not an attorney

and I have not thoroughly researched the law on patent invalidity. I am relying only on instructions from Petitioner's attorneys for these legal standards.

**A. Claim Construction**

29. I have been instructed by counsel that claim construction is a matter of law for the arbiter of law to decide. I understand that in an *inter partes* review, claims are construed using the same claim construction standard that would be used to construe the claim in a civil action.

30. I understand that a patent may include two types of claims, independent claims and dependent claims. An independent claim stands alone and includes only the limitations it recites. A dependent claim can depend on an independent claim or another dependent claim. I understand that a dependent claim includes all the limitations that it recites in addition to the limitations recited in the claim from which it depends.

**B. Anticipation**

31. I understand that a patent claim is anticipated when a single piece of prior art describes every element of the claimed invention, either expressly or inherently, arranged in the same way as in the claim. For inherent anticipation to be found, it is required that the missing descriptive material is necessarily present in the prior art. I understand that, for the purpose of an *inter partes* review, prior art that

anticipates a claim can include both patents and printed publications from anywhere in the world.

### **C. Obviousness**

32. I understand that a patent claim is unpatentable and invalid if the subject matter of the claim as a whole would have been obvious to a POSITA as of the time of the invention at issue. My understanding of a POSITA is set forth below. I understand that the following factors must be evaluated to determine whether the claimed subject matter is obvious: (1) the scope and content of the prior art; (2) the difference or differences, if any, between each claim of the patent and the prior art; and (3) the level of ordinary skill in the art at the time the patent was filed. Unlike anticipation, which allows consideration of only one item of prior art, I understand that obviousness may be shown by considering more than one item of prior art. Moreover, I have been informed and I understand that the so-called objective indicia of non-obviousness, also known as “secondary considerations,” are also to be considered when assessing obviousness. These include: (1) commercial success; (2) long-felt but unresolved needs; (3) copying of the invention by others in the field; (4) initial expressions of disbelief by experts in the field; (5) failure of others to solve the problem that the inventor solved; and (6) unexpected results. I also understand that evidence of objective indicia of non-obviousness must be commensurate in scope with the claimed subject matter.

### **III. The '316 Patent**

33. The '316 patent is titled “Integrated Shopping Within Long-Form Entertainment.”

34. The '316 patent lists inventors Charley Pavlosky, Traci Gilland, Tony Fannin, Eric Wickens, and John Patrick Wooton.

35. The '316 patent was filed as U.S. Patent Application No. 15/855,965 on December 27, 2017, and issued on September 28, 2021.

#### **A. Priority Date**

36. The '316 patent claims priority to U.S. Provisional Application No. 62/439,520, filed on December 28, 2016. I have assumed this date for purposes of my analysis, but do not offer an opinion whether or not the '520 provisional supports December 28, 2016, as the priority date.

#### **B. Specification**

37. The specification assumes familiarity with “television and movies as a form of long-form entertainment,” “commercials ... directly advertising between programming segments on television,” and “online entertainment services.” EX1001, 1:19-20, 1:27-29, 1:67-2:2. The specification acknowledges that the “online streaming and advertising service[s]” already existed at the time of alleged invention including “NETFLIX®, HULU®, HBO®, and AMAZON®.” EX1001, 2:10-13, 1:37-42. The specification also acknowledges the existence of “traditional

television guide layers, channel navigation layers, TV show or movie information layers, or closed captioning layers.” EX1001, 14:13-17. The ’316 then assumes that existing programming languages may be used to create the interactive video layers described therein but lacks any disclosure for implementing the same. EX1001, 11:67-12:5, 13:46-52. The specification does not, however, disclose or purport to disclose new or improved means of generating layers, recognizing objects in video media, or incorporating social media functionality.

38. Rather, the ’316 patent purports to use this existing technology to “integrate[] a social media service within the entertainment without interruption of the entertainment,” “include a shopping layer that ... is immune to ad blocking,” and “create a unique model for online advertising within the long-form entertainment.” EX1001, 2:13-21, 2:26-27. But the prior art already provided these benefit. *See, e.g.*, EX1006, 25:47-54 (describing “a software plug-in application for a social networking site ... which enables the user to add the video object to the user’s social network account without leaving the portal ... through which the user was viewing the video”), 22:57-61 (describing a system for “the auction of advertising rights to a video,” including movies); *see also* EX1010, 8:55-59 (“In addition, advertisements placed within the video stream as described above may be insulated from ad blocking software.”).

**C. Prosecution History of the '316 Patent (EX1004)**

39. During prosecution, Applicant made several amendments to the claims in response to examiner rejections. EX1004, 124, 166, 237, 311, 343. The examiner rejected the claims under 35 U.S.C. § 103 as unpatentable in view of U.S. Pat. Pub. No. 2012/0167145 (“Incorvia”), U.S. Pat. Pub. No. 2012/0159327 (“Law”), U.S. Pat. No. 10,440,436 (“Taylor”), and U.S. Pat. Pub. No. 2005/0262542 (“DeWeese”). EX1004, 104-14, 140-51, 178-90, 215-29, 260-75. None of those references are relied upon in the Petitioner’s *inter partes* review petition.

40. The original claims did not recite most of the limitations regarding the social media portion beyond that the user interacts with a guest. *Id.*, 63, 65, 66. Over a series of rejections and amendments, the applicant added the current limitations.

41. On October 30, 2018, the examiner rejected the pending claims in view of Incorporvia, which discloses tracking objects within a video, and Law, which discloses social media interaction of television content. *Id.*, 103-13. The applicant responded by amending the claims to add that the social media guest “is also a viewer of the video media.” *Id.*, 130-33. The examiner issued another rejection, noting that Law explicitly discloses such a guest. *Id.*, 140-151. The applicant then added the limitation that the guest interaction is displayed in the second video layer, arguing that Law does not display the interaction. *Id.*, 167-71. The examiner disagreed, pointing to Law’s disclosure of instant messages that are overlaid on the television

content. *Id.*, 178-90. The applicant further added the limitation that the overlay include past interaction between a plurality of users, arguing that Law only discloses a single instant message as an interactive overlay. *Id.*, 196-99, 201-2.

42. The examiner rejected the amended claims in view of Incorvia, Law, and Taylor, noting that the latter discloses social media interaction overlaid on a video where past interaction with other users is included. *Id.*, 215-29. In response, the applicant then added that the limitation reciting toggling between different conversations within the overlay. *Id.*, 239-42. The examiner again rejected the claims, this time in further view of DeWeese, which discloses a chat overlay on a television program that allows toggling between different conversations. *Id.*, 261-75.

43. On April 27, 2021, the applicant and examiner conducted an interview where they agreed on amendments to overcome the prior art. *Id.*, 308. The amendments recited that the user and guest are each provided with the video and overlays via their respective media playback devices. *Id.*, 312-16. The applicant argued that Law only discloses two users watching a video on the same television, though commenting on it via their own devices. *Id.*, 318. The applicant also argued that DeWeese only allows users to chat above television programs they are watching, but fails to disclose that each user views the video and overlays on their respective devices. *Id.*, 318. These claims were allowed after the applicant added clarifying

language that the user is included in the conversation to overcome a Section 112 rejection. *Id.*, 343-44, 352-56.

44. Accordingly, as noted by the examiner, the social media interactions recited in the claims were well-known in the art. The examiner only allowed the claims once they were amended to specify that each user views the video and overlays on their own respective devices. However, as discussed below, this limitation was disclosed in Ohliger.

**D. Level of Ordinary Skill in the Art**

45. I understand that a person of ordinary skill in the art (“POSITA”) is a hypothetical person who is presumed to be aware of all pertinent art, thinks along conventional wisdom in the art, and is a person of ordinary creativity—not an automaton. In deciding the level of ordinary skill, I understand that the following factors may be considered:

- The levels of education and experience of persons working in the field;
- The types of problems encountered in the field; and
- The sophistication of the technology.

46. I understand that the asserted claims of the asserted patents must be evaluated from the perspective of a POSITA. In my opinion, the relevant art for this patent relates to interactive multimedia systems including video and audio processing, *i.e.*, applications for interactive video. I understand that the relevant

point in time for determining the qualifications of a person of ordinary skill in the state-of-the-art is the time of the alleged invention, which I assume to be the earliest effective filing date for the patent. Here, I understand that the earliest alleged priority date is December 27, 2016.

47. In my opinion, a person of ordinary skill in the art at the time of the '316 patent ("POSITA") had at least a Bachelor of Science in electrical engineering, computer engineering, or similar fields and at least two years of practical experience in the field of multimedia systems including video processing and audio processing. This level of skill is approximate, and more experience would compensate for less formal education, and vice versa.

48. I meet these criteria now and met them at the time of the alleged invention. I have applied this level of skill in my analysis. My opinions would not change if a slightly higher or lower level of ordinary skill applied.

**E. Claim Construction**

49. I have been instructed by Petitioner to perform my technical analysis of the disclosures of the prior art by relying on the following constructions presented in district court. Otherwise, I have applied the plain and ordinary meaning of all other claim terms unless noted otherwise below.

| <b>Term</b> | <b>Petitioner's Construction</b> | <b>PO's Construction</b> |
|-------------|----------------------------------|--------------------------|
|-------------|----------------------------------|--------------------------|

|                                                                                                                                                                                                                                                          |                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                              |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| “a plurality of ... audio objects in one or more frames of the video media” (claim 1)                                                                                                                                                                    | Indefinite. Claim term is nonsensical because audio objects cannot be “in” frames. EX1016, 3-6.                                                                                   | Plain meaning. The claim uses “frames as a temporal marker” to signify the “objects in question may persists through only a portion of the video media.” EX1017, 4-5.                                                                                                                                                                                                                                                                        |
| For this term, I relied on the prior art’s disclosure of “a plurality of visual ... object in one or more frames of the video,” which means the parties’ dispute regarding “audio objects” is inapplicable to my mapping of the prior art to the claims. |                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| “correspondingly form[ing] one or more items, services, or songs of interest” and “the one or more items, services, or songs in the video media” (claim 1)                                                                                               | Indefinite. Claim term is contradictory because “items, services, or songs” cannot be both “in” the received “video media” and “form[ed]” while processing the same. EX1016, 6-8. | Plain and ordinary meaning. Claim 1 distinguishes between “items, services, or songs” and “items, services, or songs of interest”; “[E]ven if there were ‘items, services, or songs’ in the video media at the time it was received, not all of them would be ‘of interest.’ ... [I]t is only when product information is associated with objects that ‘one or more items, services, or songs <i>of interest</i> ’ are formed.” EX1017 at 7. |
| For this term, I applied Patent Owner’s construction.                                                                                                                                                                                                    |                                                                                                                                                                                   |                                                                                                                                                                                                                                                                                                                                                                                                                                              |
| “items, services, or songs presented in a video media” and “said first interactive layer contains the items, services, or songs of interest” (claim 18)                                                                                                  | Indefinite. Claim term is contradictory because “items, services, or songs” cannot be both “presented in a video media” and “contain[ed]”                                         | Plain and ordinary meaning. Claim 18 “does not require that the items, services or songs are <i>contained</i> in the video media.” Instead, it claims ‘items, services or songs                                                                                                                                                                                                                                                              |

|                                                                              |                                                                                       |                                                                                                                                                                                                                                                               |
|------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                              | “in the first interactive layer.” EX1016, 6-8.                                        | <b><i>presented</i></b> in the video media.” In addition, “items” are distinguishable from “items ... of interest,” and only the “items ... of interest” are “contain[ed]” in the first interactive video layer. EX1017, 8.                                   |
| For this term, I applied Patent Owner’s construction.                        |                                                                                       |                                                                                                                                                                                                                                                               |
| Certain limitations of independent claims 1, 17, and 18.                     | Certain limitations must be performed in order. EX1016, 9-12.                         | Plain and ordinary meaning. “Each of claims 1, 17, and 18 ... should not be construed to require any particular order of method steps.” EX1017, 10-11.                                                                                                        |
| For this term, I applied both Petitioner’s and Patent Owner’s constructions. |                                                                                       |                                                                                                                                                                                                                                                               |
| “processor” (claim 17)                                                       | At least one processor performs all claimed steps.                                    | Plain and ordinary meaning. “[A] POSITA would understand that the claims only require that ‘one or more’ processors collectively perform all claimed steps and that ‘one or more’ applications collectively contain all claimed software modules. EX1017, 16. |
| “application” (claim 18)                                                     | At least one application contains all claimed software modules.                       |                                                                                                                                                                                                                                                               |
| For this term, I applied both Petitioner’s and Patent Owner’s constructions. |                                                                                       |                                                                                                                                                                                                                                                               |
| “software module” terms of claim 18                                          | Means-plus-function terms; indefinite for lack of sufficient corresponding structure. | Plain and ordinary meaning. These “software module” limitations are not subject to 35 U.S.C. § 112(f) because “[t]he                                                                                                                                          |

|                                                       |  |                                                                                                                   |
|-------------------------------------------------------|--|-------------------------------------------------------------------------------------------------------------------|
|                                                       |  | structure of the software modules in claim 18 is explicitly described in the preamble of claim 18[.]” EX1017, 16. |
| For this term, I applied Patent Owner’s construction. |  |                                                                                                                   |

50. I reserve the right to provide additional opinions concerning claim construction or the application of certain claim constructions to the prior art, as appropriate, and to respond to any particular claim construction-related argument advanced by Patent Owner and/or its expert.

51. In my opinion, no express constructions are required to find the ’316 patent claims invalid. I understand that in parallel district court litigation, Petitioner has asserted that certain claim terms in independent claims 1 and 18 (but not independent claim 17) are indefinite, as a POSITA would be unable to determine the outer boundaries for those terms with reasonable certainty. EX1016. In my opinion, these arguments do not affect my opinions as expressed in this declaration, as the prior art discloses the limitations whether or not the outer bounds of the term are definite. *Mylan Pharms. Inc. v. Horizon Pharma USA*, IPR2017-01995, Paper 72 at 5-6 (Mar. 28, 2019) (“[W]e need not necessarily determine the outer boundaries of a claim in reaching a conclusion as to whether the asserted prior art anticipates a claim or renders it obvious.”).

52. I understand that the parties also dispute in the district court litigation (1) whether certain limitations of independent claims 1, 17, and 18 must be performed in order, (2) whether at least one processor must perform all steps of claim 17, (3) whether at least one application must contain all software modules of claim 18, and (4) interpretation of “software module” terms in claim 18. EX1016; EX1017. In my opinion, the prior art discloses or renders obvious the challenged claims under either interpretation of those limitations.

53. As described in more detail in the remainder of this declaration, the prior art discloses or renders obvious the challenged claims under any reasonable potential claim interpretation.

#### **F. Challenged Claims**

54. I understand that claims 1-19 are at issue in Petitioner’s petition for *inter partes* review. They are reproduced below for reference.

| <b>Claim</b> | <b>Limitation</b>                                                                                                    |
|--------------|----------------------------------------------------------------------------------------------------------------------|
| 1[pre]       | A method for providing information on items, services, or songs in a video media comprising:                         |
| 1[a]         | a) receiving the video media having a plurality of visual or audio objects in one or more frames of the video media; |
| 1[b]         | b) processing the video media to generate a first interactive video layer, wherein the processing includes:          |
| 1[b][i]      | i. identifying one or more of the plurality of visual or audio objects within the video media,                       |

| <b>Claim</b> | <b>Limitation</b>                                                                                                                                                                                                                                                                                                                           |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1[b][ii]     | ii. associating product information to the one or more of the plurality of visual or audio objects to correspondingly form one or more items, services, or songs of interest;                                                                                                                                                               |
| 1[b][iii]    | iii. tracking the one or more items, services, or songs of interest through the one or more frames of the video media; and                                                                                                                                                                                                                  |
| 1[b][iv]     | iv. assigning a selectable visual indicator for each of the one or more items, services, or songs of interest, wherein the first interactive video layer is configured to display the selectable visual indicator;                                                                                                                          |
| 1[c][i]      | c) generating a video data comprising the video media, the first interactive video layer, and a second interactive video layer,                                                                                                                                                                                                             |
| 1[c][ii]     | wherein the video data is configured to be provided to a user via a first media playback device and to at least one guest via a second media playback device,                                                                                                                                                                               |
| 1[c][iii]    | wherein the second interactive video layer comprises a social media portion controllable by the user via the first media playback device, so as to interact with the at least one guest that is also a viewer of the video media through an interface displayed in the second interactive video layer via the second media playback device, |
| 1[c][iv]     | wherein the second interactive video layer displays an interaction between the user and the at least one guest, and past interaction activity between a plurality of users in the social media portion, and                                                                                                                                 |
| 1[c][v]      | wherein the user is capable of maintaining and toggling between a plurality of different conversations displayable in the social media portion in the second interactive video layer as an overlay to the video media displayed by the first media playback device,                                                                         |
| 1[c][vi]     | wherein the plurality of users either includes the user or does not include the user; and                                                                                                                                                                                                                                                   |
| 1[d]         | d) distributing information associated with the one or more items, services, or songs in the video media to the user.                                                                                                                                                                                                                       |

| <b>Claim</b> | <b>Limitation</b>                                                                                                                                                                                                            |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 2[pre]       | The method of claim 1, further comprising:                                                                                                                                                                                   |
| 2[a][i]      | enabling the user to present the video data with or without the first interactive video layer being turned on or off based on selection by the user to view or not view the first interactive video layer, respectively; and |
| 2[a][ii]     | enabling the user to present the video data with or without the second interactive video layer being turned on or off based on selection by the user to view or not view the second interactive video layer, respectively.   |
| 3[pre]       | The method of claim 2, further comprising:                                                                                                                                                                                   |
| 3[a][i]      | displaying the video media and the first interactive video layer to the user; and                                                                                                                                            |
| 3[a][ii]     | receiving a selection of a given selectable visual indicator of a given item, service, or song of interest from the user to display marketing information of the given item, song, service, or song of interest to the user. |
| 4[pre]       | The method of claim 3, further comprising                                                                                                                                                                                    |
| 4[a]         | enabling the user to purchase the given item, song, or service of interest from the marketing information.                                                                                                                   |
| 5[pre]       | The method of claim 4,                                                                                                                                                                                                       |
| 5[a]         | wherein the user is enabled to purchase while the video media is being displayed to the user.                                                                                                                                |
| 6[pre]       | To the extent the preamble is limiting, The method of claim 1, further comprising                                                                                                                                            |
| 6[a]         | copying a timed portion of the video data to a clip list based on a selection of an item, service, or song of interest by the user.                                                                                          |
| 7[pre]       | The method of claim 1, further comprising                                                                                                                                                                                    |

| <b>Claim</b> | <b>Limitation</b>                                                                                                                                                                                                                             |
|--------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7[a]         | collecting data based on actions taken by the user.                                                                                                                                                                                           |
| 8[pre]       | The method of claim 1, further comprising:                                                                                                                                                                                                    |
| 8[a]         | removing the product information from the one or more of the plurality of visual or audio objects; and associating different product information with the one or more of the plurality of visual or audio objects.                            |
| 9[pre]       | The method of claim 1, further comprising                                                                                                                                                                                                     |
| 9[a]         | providing a third interactive video layer having one or more visual indicators to the server for distribution with the video media to the user.                                                                                               |
| 10[pre]      | The method of claim 9,                                                                                                                                                                                                                        |
| 10[a]        | wherein the one or more visual indicators include a link for one or more of a second video media, a side room digital layer, a promotional video, a game, and/or merchandise.                                                                 |
| 11[pre]      | The method of claim 1,                                                                                                                                                                                                                        |
| 11[a]        | wherein the video data includes a third interactive video layer, wherein the third interactive video layer comprises one or more visual indicators.                                                                                           |
| 12[pre]      | The method of claim 1,                                                                                                                                                                                                                        |
| 12[a]        | wherein the video media includes one of a television episode or a movie.                                                                                                                                                                      |
| 13[pre]      | The method of claim 1,                                                                                                                                                                                                                        |
| 13[a]        | wherein the identifying one or more of the plurality of visual or audio objects is performed manually through input of spatial and/or temporal data for each of the one or more items, services, or songs of interest within the video media. |
| 14[pre]      | The method of claim 1,                                                                                                                                                                                                                        |

| <b>Claim</b> | <b>Limitation</b>                                                                                                                                                                                                                                     |
|--------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 14[a]        | wherein identifying one or more of the plurality of visual or audio objects is performed automatically by a computer program for visual and/or audio recognition and mapping.                                                                         |
| 15[pre]      | The method of claim 1,                                                                                                                                                                                                                                |
| 15[a]        | wherein the first and second interactive layers are unaffected by ad blocking software.                                                                                                                                                               |
| 16[pre]      | The method of claim 1,                                                                                                                                                                                                                                |
| 16[a][i]     | wherein the interaction and the past interaction activity comprise a public or private conversation between the user and the at least one guest,                                                                                                      |
| 16[a][ii]    | wherein the plurality of different conversations comprises at least a public conversation which enables all viewers of the video media to join and a private conversation which enables a permitted subset of all viewers of the video media to join. |
| 17[pre]      | A non-transitory computer readable media encoded with instructions that, when executed by a processor, cause the processor to carry out at least the following steps:                                                                                 |
| 17[a]        | (a) processing video media to generate a first interactive video layer, wherein the processing includes:                                                                                                                                              |
| 17[a][i]     | i. identifying one or more of the plurality of visual or audio objects within the video media;                                                                                                                                                        |
| 17[a][ii]    | ii. associating product information to the one or more of the plurality of visual or audio objects to correspondingly form one or more items, services, or songs of interest;                                                                         |
| 17[a][iii]   | iii. tracking the one or more items, services, or songs of interest through the one or more frames of the video media; and                                                                                                                            |
| 17[a][iv]    | iv. assigning a selectable visual indicator for each of the one or more items, services, or songs of interest, wherein the first interactive video layer is configured to display the selectable visual indicator; and                                |

| <b>Claim</b> | <b>Limitation</b>                                                                                                                                                                                                                                                                                                                                                                                              |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 17[b][i]     | (b) generating a video data comprising the video media, the first interactive video layer, and a second interactive video layer,                                                                                                                                                                                                                                                                               |
| 17[b][ii]    | wherein the video data is configured to be provided to a user via a first media playback device and to at least one guest via a second media playback device,                                                                                                                                                                                                                                                  |
| 17[b][iii]   | wherein the second interactive video layer comprises a social media portion controllable by the user via the first media playback device, so as to interact with the at least one guest that is also a viewer of the video media through an interface displayed in the second interactive video layer via the second media playback device,                                                                    |
| 17[b][iv]    | wherein the second interactive video layer displays an interaction between the user and the at least one guest, and                                                                                                                                                                                                                                                                                            |
| 17[b][v]     | past interaction activity between a plurality of users in the social media portion, and                                                                                                                                                                                                                                                                                                                        |
| 17[b][vi]    | wherein the user is capable of maintaining and toggling between a plurality of different conversations displayable in the social media portion in the second interactive video layer as an overlay to the video media displayed by the first media playback device,                                                                                                                                            |
| 17[b][vii]   | wherein the plurality of users either includes the user or does not include the user.                                                                                                                                                                                                                                                                                                                          |
| 18[pre]      | A computer-implemented system comprising: a digital processing device comprising: at least one processor, an operating system configured to perform executable instructions, a memory, and a computer program including instructions executable by the digital processing device to create an application for providing information on items, services, or songs in a video media, the application comprising: |
| 18[a]        | a) a software module for receiving a request for the generation of information regarding items, services, or songs presented in a video media from a first processing device of a user;                                                                                                                                                                                                                        |

| <b>Claim</b> | <b>Limitation</b>                                                                                                                                                                                                                                                                                                            |
|--------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 18[b]        | b) a software module for linking the items, services, or songs with marketing and/or purchasing information;                                                                                                                                                                                                                 |
| 18[c]        | c) a software module for the generation of a first interactive layer to display with the video media, wherein said first interactive layer contains the items, services, or songs of interest linked to the marketing and/or purchasing information;                                                                         |
| 18[d][i]     | d) a software module for the generation of a second interactive layer to display with the video media and optionally the first interactive layer,                                                                                                                                                                            |
| 18[d][ii]    | wherein the video media, the first interactive layer, and the second interactive layer are configured to be provided to user via the first processing device and to at least one guest via a second processing device,                                                                                                       |
| 18[d][iii]   | wherein the second interactive layer contains a social media portion controllable by the user via the first processing device, so as to interact with the at least one guest that is also a viewer of the video media through an interface displayed in the second interactive video layer via the second processing device, |
| 18[d][iv]    | wherein an interaction between the user and the at least one guest, and past interaction activity between a plurality of users in the social media portion, are displayed in the interface in the second interactive video layer, and                                                                                        |
| 18[d][vi]    | wherein the user is capable of maintaining and toggling between a plurality of different conversations displayable in the social media portion in the second interactive video layer as an overlay to the video media displayed by the first processing device,                                                              |
| 18[d][vii]   | wherein the plurality of users either includes the user or does not include the user; and                                                                                                                                                                                                                                    |
| 18[e]        | e) a software module for transmitting the first and second interactive layers to the first processing device of the user.                                                                                                                                                                                                    |

| Claim     | Limitation                                                                                                                                                                                                                                            |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 19[pre]   | The computer-implemented system of claim 18,                                                                                                                                                                                                          |
| 19[a][i]  | wherein the interaction and the past interaction activity comprise a public or private conversation between the user and the at least one guest,                                                                                                      |
| 19[a][ii] | wherein the plurality of different conversations comprises at least a public conversation which enables all viewers of the video media to join and a private conversation which enables a permitted subset of all viewers of the video media to join. |

#### IV. State of the Art

55. The '316 patent simply combines well-established technologies that were already known and used according to their conventional purposes by at least December 28, 2016. These include interactive video layers and overlays, object tracking in video content, and social media chat functionalities. Further, the patent does not describe particular challenges or solutions in putting these well-known techniques together and does not offer any special benefit or synergy from putting these existing technologies together.

56. ***Video Layers/Overlays:*** Superimposing interactive layers on video has been a standard technique for decades. Interactive program guides emerged in the 1980s, and by the 1990s, it was common practice to overlay such guides on television programs. These weren't just simple overlays either—the guides in the 1990s could embed video content and allow integration within the guide itself. For instance, U.S. Pat. No. 6,388,714, filed April 11, 1997 and issued May 14, 2002,

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

clearly describes both embedding broadcast television into a program area 726, as well as displaying parts of the guide over the video. *See, e.g.*, U.S. Pat. No. 6,388,714, Fig. 12A, 13C. The interactivity wasn't limited to program schedules—it enabled linking to services and products related to the underlying video, such as ordering merchandise or subscription services associated with an NFL game being watched. *See, e.g.*, U.S. Pat. No. 6,388,714, FIG. 17C.

57. ***Tracking Objects in Video:*** Tracking objects in video has been a well-established field for decades. At its most basic level, this involves calculating a 2D cross-correlation between a video frame and an object representation—mathematical techniques that date back to the 18th century. More sophisticated approaches like kernel-based tracking, contour tracking, Kalman filtering, and Particle filtering have been available since the 1960s. By the time of the '316 patent's priority date, sophisticated tracking algorithms were freely available as open source software that anyone could download and use. *See, e.g.*, EX1018.

58. ***Social Media and Chat:*** modern computer-based social media and chat systems developed through distinct phases, each building on previous innovations. The foundation was laid in the 1970s and early 1980s with bulletin board systems (BBS) and early email protocols, which operated asynchronously. The late 1980s and 1990s brought real-time communication—America Online's instant messaging brought one-to-one messaging to mainstream users. Social networking gained

momentum in the 1990s with SixDegrees launching in 1997 as what's often considered the first social media site, followed by Friendster (2002), Myspace (2003), YouTube (2005), and Facebook (2006). Importantly, social media platforms integrated chat and commenting from the very beginning. Facebook's commenting system from 2006 became ubiquitous through social plugins (U.S. Patent No. 9,311,678), YouTube implemented commenting in 2005 for video discussions, and later platforms like Meerkat and Periscope (2015) built mobile live streaming with real-time comments, while Twitch (2011) centered its entire platform around live chat accompanying gaming streams.

## V. Grounds

| Ground | Basis | Reference                                       | Claims |
|--------|-------|-------------------------------------------------|--------|
| 1      | §103  | Ohliger (U.S. No. 11,432,046)                   | 1-19   |
| 2      | §103  | Ohliger and Sakhartov (U.S. No. 8,166,500)      | 1-19   |
| 3      | §103  | Ohliger and Kaiserlian (U.S. No. 8,332,477)     | 1-19   |
| 4      | §103  | Ohliger and Tapissier (E.P. No. 1,425,918)      | 1-19   |
| 5      | §103  | Ohliger, Sakhartov, and Kaiserlian or Tapissier | 1-19   |

### A. Ground 1: Ohliger

#### 1. Background

59. Ohliger qualifies as prior art under at least AIA 35 U.S.C. § 102(a), having been filed June 13, 2016, which predates the earliest effective filing date of the '316 patent. Ohliger discloses all limitations of the '316 patent claims, but I understand Petitioner presents Ohliger as an obviousness reference in the event

Patent Owner argues there are differences between Ohliger and the '316 patent claims. My opinion is that Ohliger at least renders all of the challenged claims obvious.

60. Ohliger describes a comprehensive system for “identifying objects in still images or in video” and making them “interactive in the image or in every frame of the video that they appear” through “an interactive layer” to connect viewers with “e-commerce sites where the object may be purchased.” EX1006, Abstract, 8:58-9:2. Figure 18,<sup>1</sup> provides a clear example of such an “interactive layer” that includes “outline around the interactive video objects” (shown as parallelograms) and displays additional information about selected objects. EX1006, 16:50-54, 19:30-37, 23:62-24:1, 24:12-20.

---

<sup>1</sup> EX1011 and EX1012 contain the figures for the Ohliger patent as provided to the Patent Office as “Supplemental Content” on the same day as the original Ohliger application and with the exact same disclosure as the application, but with higher resolution. EX1015. For ease of review, the higher resolution versions of Ohliger’s figures are used in this declaration.

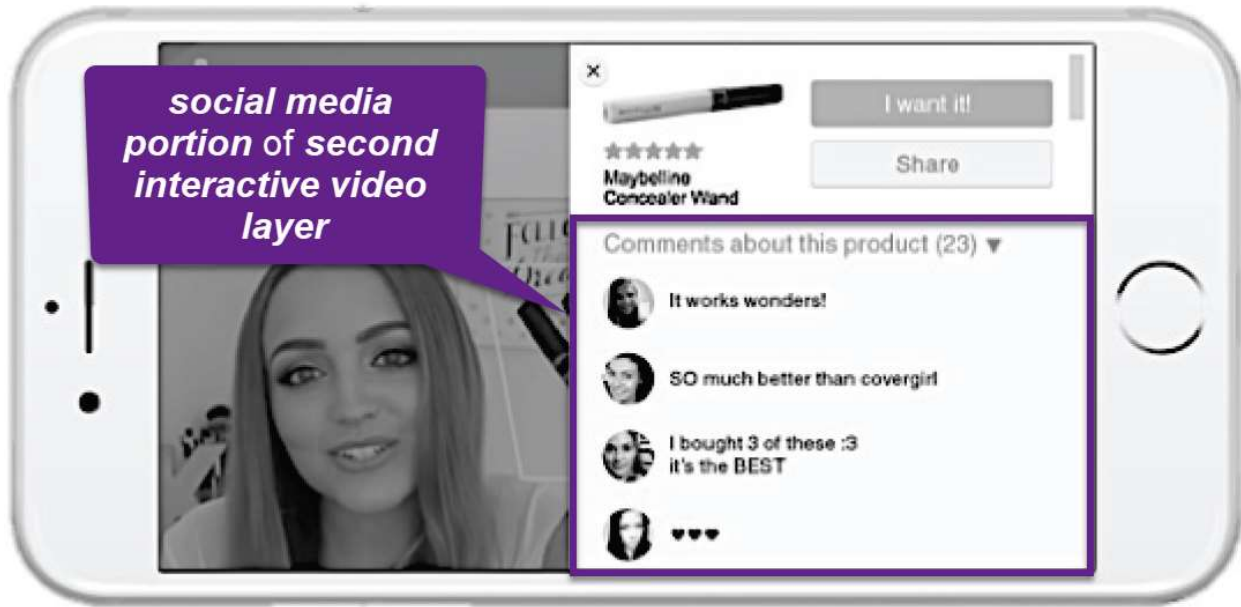


2

61. Ohliger also discloses multiple integrations with social media, including allowing users to post videos on social media (Fig. 15), leave general comments relevant to the video (Fig. 16), or leave specific comments about individual video objects available for purchase (Fig. 26-27). *See, e.g.,* EX1006, Fig. 26:

---

<sup>2</sup> All figures contain color annotations.



62. The system achieves these overlays through what Ohliger calls its “video object hyperlink exchange” (VOHE) server 300. EX1006, 7:36-38, 15:27-28, 19:13-18, 22:4-7. Within VOHE server 300, the video database 108 contains the “raw video content.” EX1006, 16:19-21. The “video content analysis application 106” then processes this “raw video content” to create the video overlays. EX1006, 16:21-24.

## 2. Analysis

### (a) Claim 1

#### (i) 1[pre]: A method for providing information on items, services, or songs in a video media comprising:

63. For purposes of my analysis, I have assumed the preamble is limiting. In my opinion, Ohliger discloses this limitation. Ohliger describes a system for overlaying price and feature information on “objects” within *video media*. These

objects encompass *items, services, or songs* that appear in the video content. Ohliger presents a method for “identifying objects in a video” and subsequently enabling viewer interaction with these identified objects. When a viewer selects a particular object, the system provides overlaid information about that object. This creates an interactive layer of data presented on top of the underlying video content. EX1006, 3:9-21 (“A system is provided for identifying objects in video either automatically by using an object database or manually by using information provided by the creator of the video, making the objects interactive in every frame of the video that they appear, allowing the creator to embed *information behind the video objects that appears when a viewer hovers over, clicks or touches the object.*”); *id.*, 2:60-65, 8:58-9:23, 11:32-37. Ohliger provides exemplary objects from a range of everyday consumer goods, including “shirt, pants, a hat, shoes, a handbag or a purse.” *Id.*, 10:3-6.

64. For each of these identified objects in the video, Ohliger’s system enables overlaying a variety of relevant pieces of information for the object. For example, Ohliger teaches that the overlaid information includes “price, features etc” about the selected object. EX1006, 24:12-16; 24:46-57 (“[T]he viewer displays an abstract of the video object in the form of an overlay ... includ[ing] descriptive information ... about the object (name, brand name, make, model, vendor, price, etc)”). This functionality is clearly depicted in Figure 5 at step 614B (annotated):

Flowchart: viewer

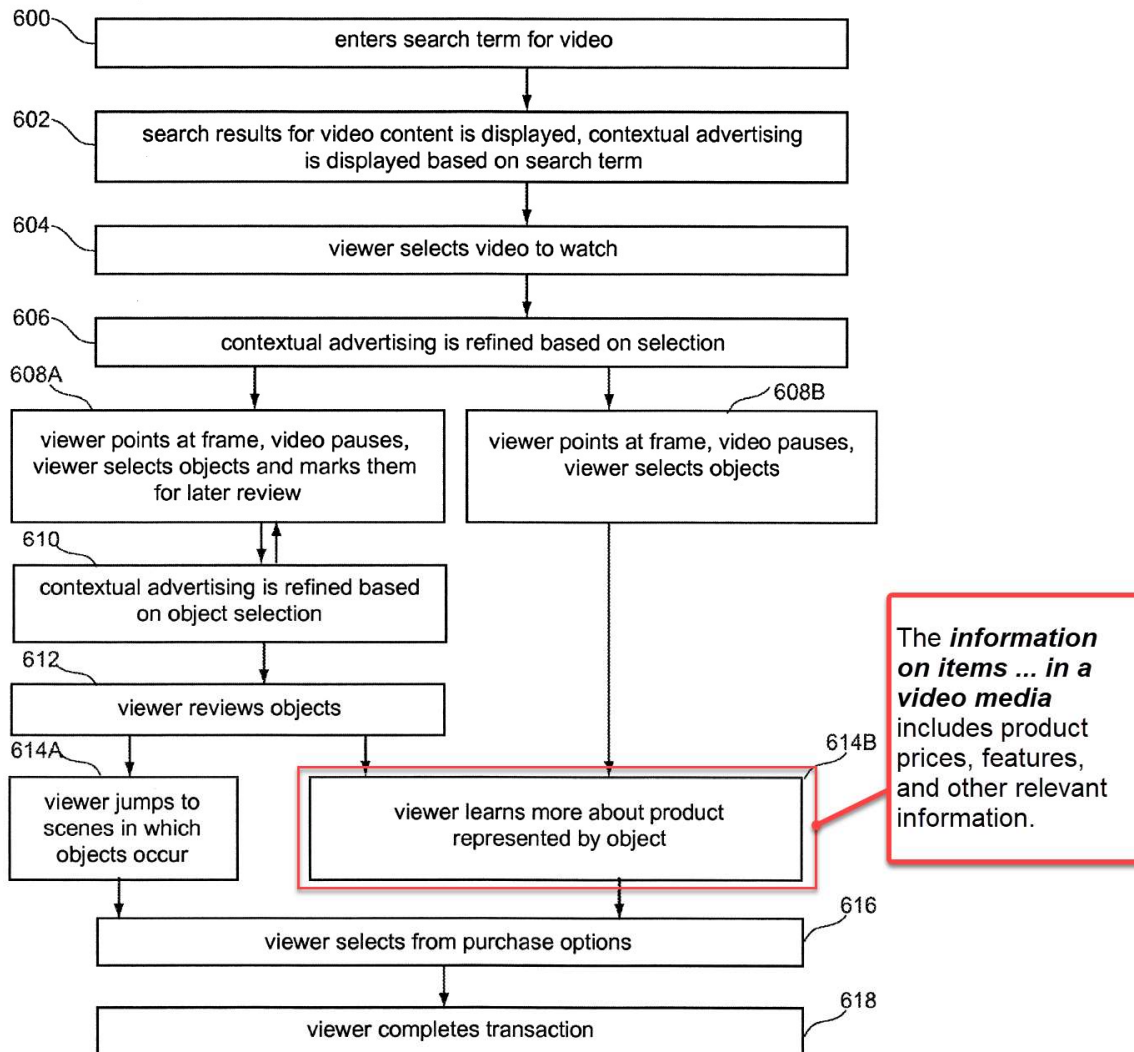


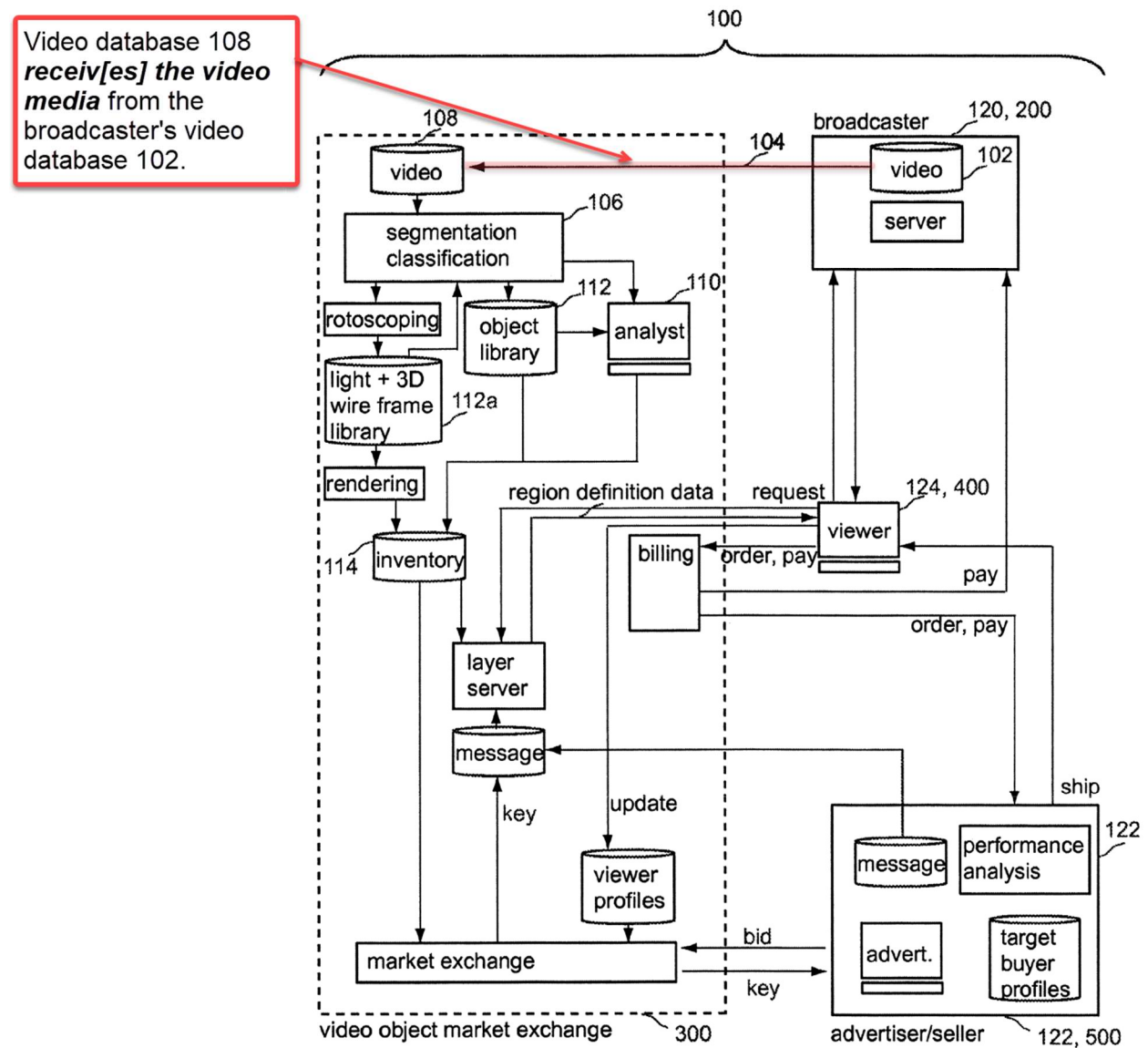
FIG. 5

- (ii) 1[a]: a) receiving the video media having a plurality of visual or audio objects in one or more frames of the video media;

65. In my opinion, Ohliger discloses this limitation. Ohliger's system features a video server that *receives* video content from a broadcaster and performs comprehensive object recognition and tracking operations. The server identifies multiple visual objects within the incoming video stream and maintains tracking of

these objects as they appear across successive frames. For this processing, the video content contains *a plurality of visual objects in multiple frames of the video*. This multi-object, multi-frame architecture enables the system to maintain persistent object identification throughout the video's duration.

66. ***“receiving the video media”***: Ohliger describes a “VOHE [video object hyperlink exchange] server 300” that serves as the central processing component for receiving video content. This server architecture includes a video database 108 that obtains video content from the broadcaster's infrastructure. As shown by the highlighted arrow in Figure 2, server 300's video database 108 receives video content from the broadcaster's video database 102. The system establishes a data flow pathway between these components for content transfer, *i.e., receiving the video media*. EX1006, 16:1-3, 15:36-44, 15:59-62 (“database 102 contains unprocessed or raw video content”), 15:66-67 (“[T]he database 102 is shown as copied to database 108”). Thus, Ohliger discloses that the video database 108 is populated with video content from the broadcaster. EX1006, 15:66-67. Therefore, video database 108 “resides on a video segmentation and classification server 300” and receives the video content from the broadcaster. EX1006, 16:1-3.

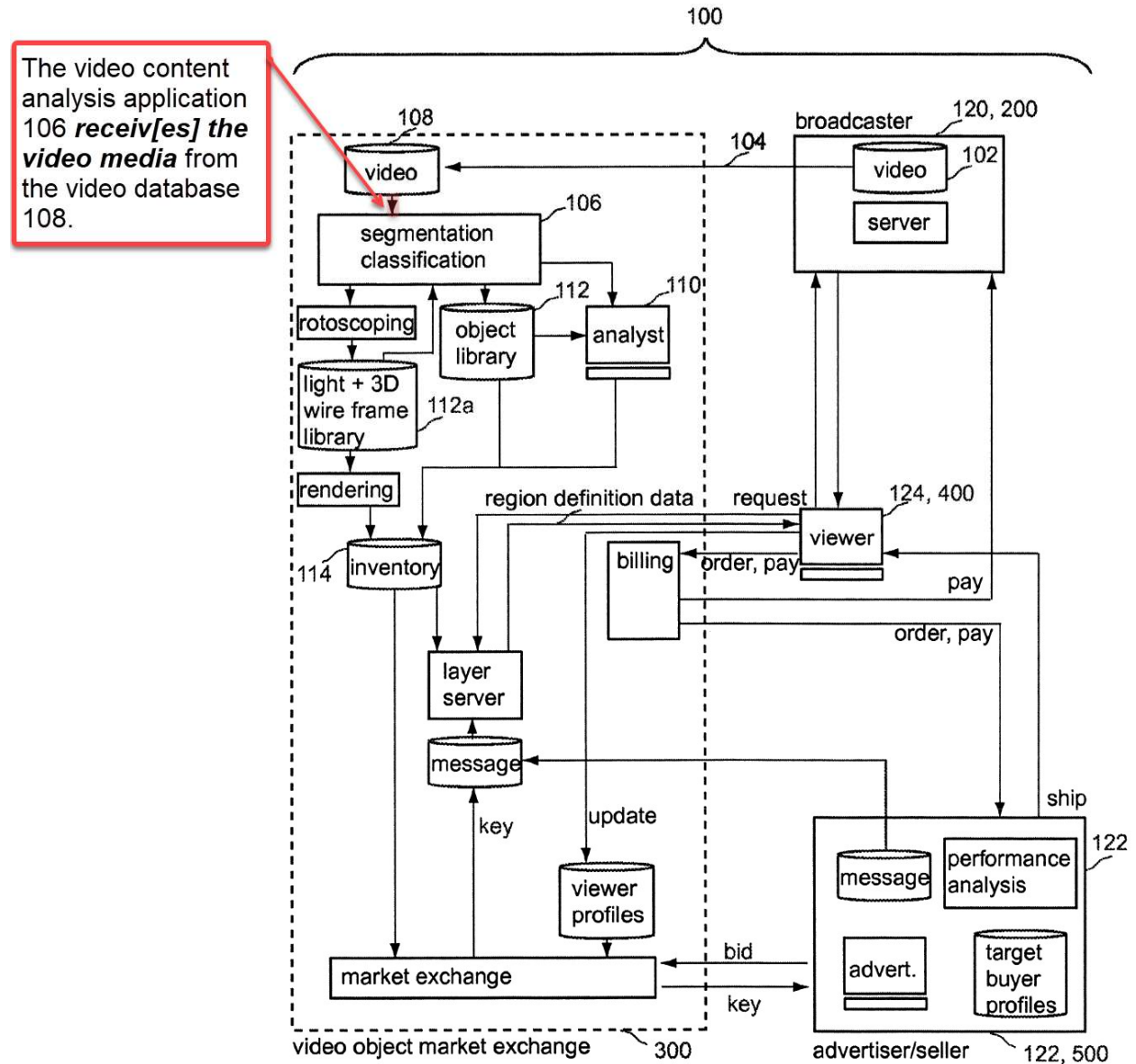


(annotated)

67. Additionally, within server 300, the “video content analysis application 106” subsequently **receives the video content** from video database 108. EX1006, 15:59-62, 16:19-24, 16:28-31, 21:34-37. Ohliger discloses this data flow for the “video content analysis application 106” as part of a robust, flexible architecture. EX1006, 16:1-9. Therefore, the video content analysis application 106 performs a

second *receiving the video media* step that is independent of the VOHE server 300

receiving the video into its video database 108.

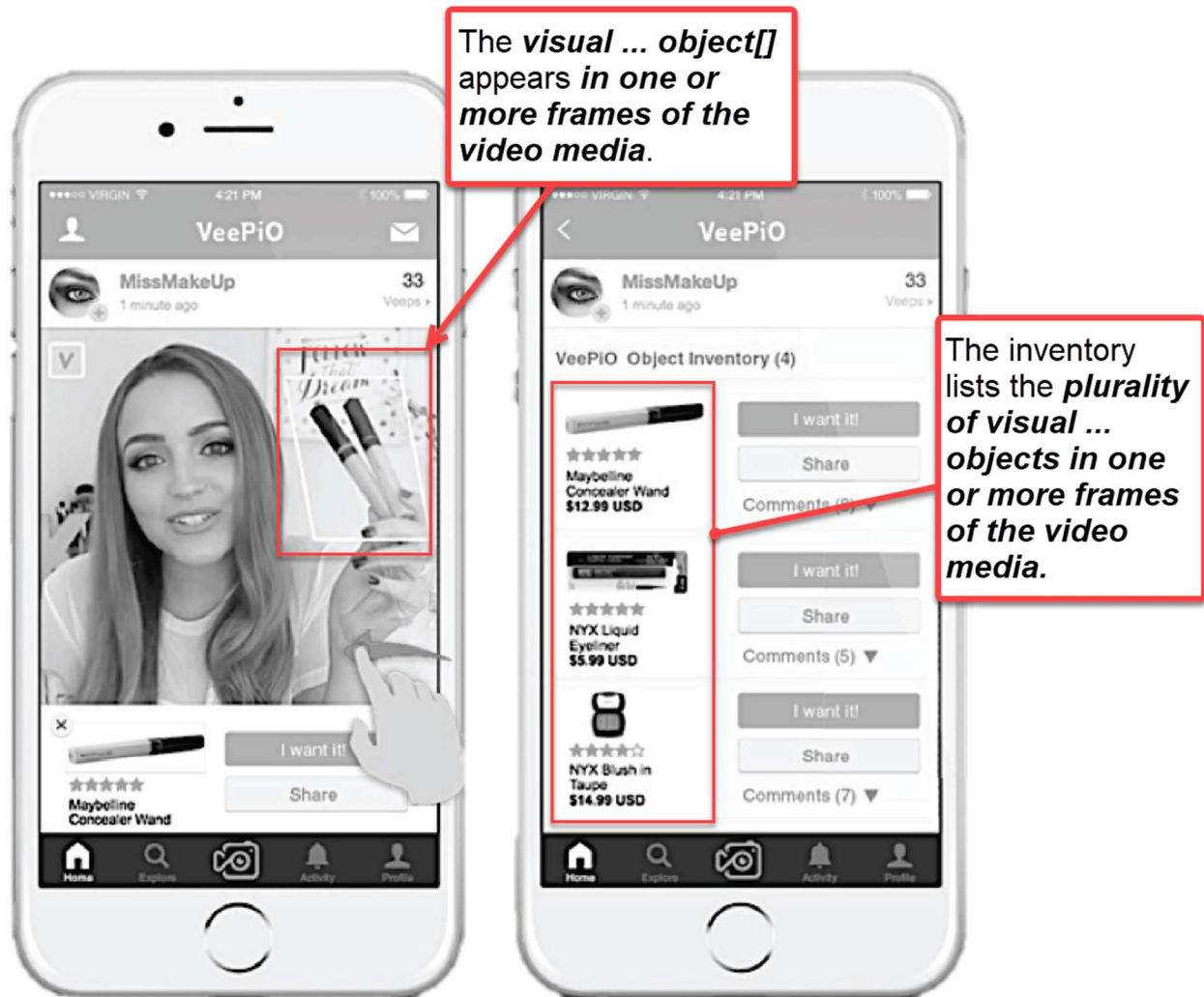


68. Additionally, Ohliger discloses a unified server architecture in the form of “a video server 200,300” that consolidates the functionality of both the broadcaster server 200 and the VOHE server 300 into a single operational unit. EX1006, 26:31-27:26. This combined server configuration performs all necessary

video processing and object identification tasks within one system. The video server 200,300 *receiv[es]* video files for processing. *Id.*, 26:36-38 (“At least one video file ... accessible to video server 200.”). This architectural approach streamlines the system by eliminating the need for separate servers while maintaining all core capabilities. Therefore, Ohliger’s combined video server 200,300 provides another independent disclosure of *receiving the video media*.

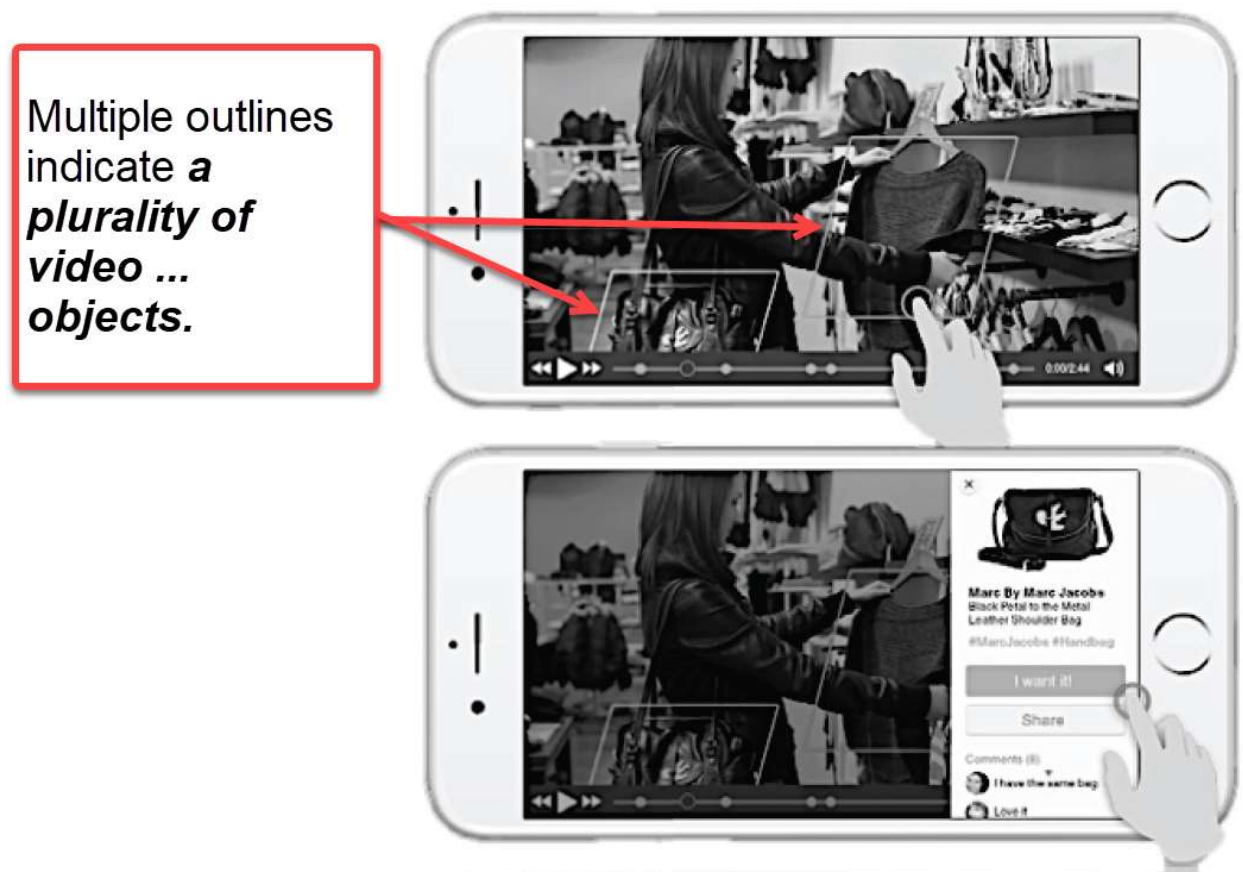
69. “*video media having a plurality of visual or audio objects*”: Ohliger teaches that the video content contains multiple “video objects,” which constitute *visual objects* within the video. Server 300, through its analysis application 106, performs comprehensive object detection by locating and classifying each video object present in the video content. EX1006, 3:37-38, 11:54-60, 16:21-28 (“[T]he video content analysis application 106 segments the raw video content to yield a list of all the video objects in a given video [and] classifies the list of video objects”). The system’s object recognition capabilities extend to tangible consumer items, with these “video objects” including articles such as “shirt, pants, a hat, shoes, a handbag or a purse which appears in the video or animation.” EX1006, 10:3-6. The practical implementation of this multi-object detection is demonstrated in Figure 27, which illustrates the server identifying multiple “Maybelline Concealer Wand” objects within the onscreen wireframe. EX1006, 40:52-54 (“multiple objects may be Veep’d for a single media source, as illustrated in FIG. 27.”). This capability confirms the

system's ability to process and track multiple distinct visual objects simultaneously within a single video source.



70. Figure 18 also shows multiple objects being identified within the wireframe overlays that the system generates. EX1006, 16:50-54. These wireframes visually delineate the boundaries of tangible objects detected within the video frames, creating distinct interactive regions around each identified item. The system employs hand cursor graphics as visual indicators to signal when these wireframe-

bounded objects become interactive elements. This visual framework establishes clear interactive zones that transform passive video content into an engaging, clickable interface where each wireframed object serves as an active hotspot for viewer interaction. As seen in Figure 18's "Object Inventory" screen, the video contains several tangible items that have been configured as such interactive elements in the video.



71. ***“in one or more frames of the video media”***: Ohliger teaches that video objects appear in multiple frames through continuous tracking mechanisms that follow objects across successive frames throughout the video’s duration. The system

maintains persistent object identification as these visual elements move through the video sequence. EX1006, 3:9-20 (“A system is provided for identifying objects in video ... making the objects interactive in every frame of the video that they appear”), 11:21-26, 10:3-6 (“The term video object as used in this specification refers to a *video frame* component”); *see also* limitation 1[b][iii] below. This frame-by-frame tracking ensures that identified objects remain consistently interactive regardless of their movement or position changes within the video content. EX1006, 16:45-49. The technical architecture supports temporal continuity, enabling the system to maintain object associations across the entire span of frames in which each object appears.

**(iii) 1[b]: b) processing the video media to generate a first interactive video layer, wherein the processing includes:**

72. In my opinion, Ohliger discloses this limitation. Ohliger’s video server *processes the video content* through a multi-stage technical workflow that transforms raw video into an interactive experience. The server first segments the video into individual frames for analysis. Within each frame, the system identifies distinct objects and associates those objects with relevant information retrieved from an object database. The server maintains temporal coherence by tracking these objects as they move from frame to frame throughout the video sequence. Following this object recognition and tracking phase, the system generates and displays

interactive bounding areas around each identified object. These bounding areas collectively form a *first interactive video layer* that overlays the original video content. This layer provides the interactive framework through which viewers can engage with the identified objects while the underlying video continues to play.

73. Figure 1 illustrates the comprehensive processing sequence for “a system for creating an interactive overlay for a video,” which Ohliger characterizes as an “interactive layer.” EX1006, 5:44-61, 8:58-9:2, 11:50-12:50. The figure delineates a systematic approach to video transformation through multiple discrete processing stages. Figure 1 and its description teach that: (1) step 700 segments the video into frames and identifies a list of objects, (2) step 710 matches object occurrences across multiple frames, (3) step 715 optionally allows video creators to assign comments or pictures to objects, (4) step 720 associates each object with respective database information, (5) step 730 creates an overlay box on each object, (6) step 745 links the overlay box with hyperlinks, and (7) steps 750-780 present the video to users with interactive overlay boxes. EX1006, 11:52-12:50. This multi-step transformation occurs within the VOHE server architecture, where the “segmentation and classification engine 106” executes these processing steps after receiving “**unprocessed** or raw video content” from the broadcaster’s database. EX1006, 15:59-62. The progression from raw content to interactive video

demonstrates how the system builds the interactive layer through a method of successive refinements of object identification, tracking, and overlay generation.

74. As discussed in further detail below, each of the claimed processing steps is met by one or more of the above steps in Figure 1. In parallel district court litigation, PO contends that claim 1 “should not be construed to require any particular order of method steps.” EX1017, 10-11. Even if PO changes positions and contends that certain steps must be performed in order, Ohliger teaches that:

75. (1) limitation 1[b][i]’s step of identifying an object in a frame is met by step 700’s frame segmentation which yields a list of objects,

76. (2) limitation 1[b][ii]’s step of associating the object with information is met by step 720’s connecting the object with corresponding database information,

77. (3) limitation 1[b][iii]’s step of tracking the object across multiple frames is met by step 730’s object tracking algorithm,

78. (4) limitation 1[b][iv]’s step of assigning an overlay to the object is met by step 730’s drawing of an overlay box or step 745’s linking of the overlay box with a hyperlink.

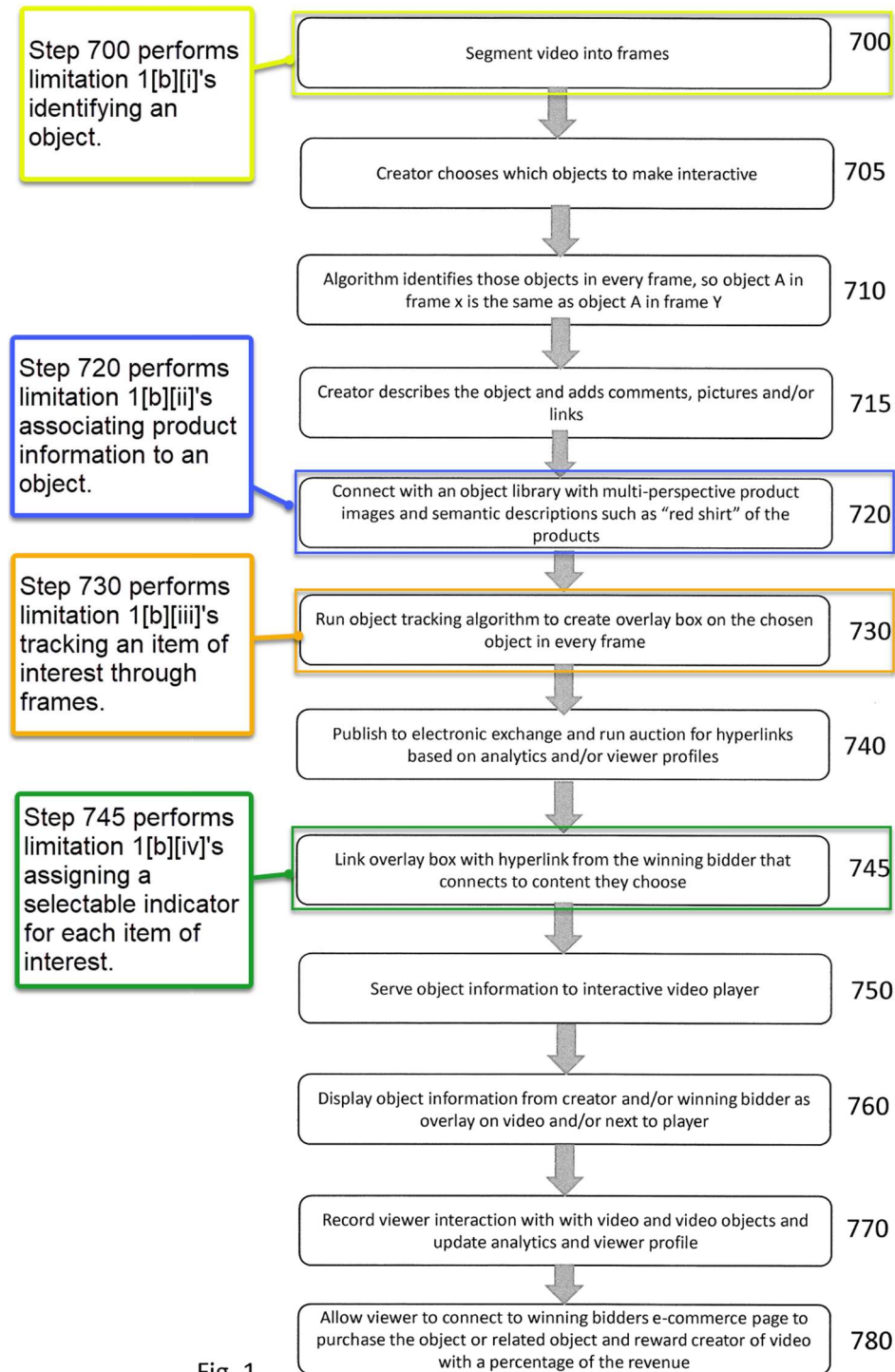


Fig. 1

79. In my opinion, other mappings of Figure 1's steps to the claimed steps are possible, as a POSITA would have understood that claimed step ii (associating information with objects) and step iii (tracking objects across frames) can be

performed in any order. The flexibility in ordering stems from the distinct nature of these operations and their varying requirements.

80. The association of information requires receiving data about the object from multiple sources, each operating under different technical constraints. Ohliger discloses several steps that individually associate different types of product information to the video objects. These include the creator's description in step 715, the object library information in step 720, and the auction-winning product hyperlink in step 740. EX1006, 12:22-24, 12:4-11, 12:29-37. Each information source requires interfacing with different software components that operate under distinct constraints with varying response times. Automated database queries typically execute faster than manual information entry from video creators. EX1006, 15:6-26.

81. Therefore, in my opinion, a POSITA would have recognized that performing the associating and tracking steps in different orders would be desirable based on the associating step's constraints. The system architecture supports bidirectional benefits between these operations. When association follows tracking, the associating step benefits from the tracking step's comprehensive object data. This provides auction bidders with complete object information based on the object's appearance throughout the entire video, including metrics such as duration of appearance, frame positioning, and relative size variations across frames. EX1006, 13:51-14:3 ("retailers/brands/affiliates/advertisers can upload a database

of images, video objects or 2D/3D wire frames to find matches in movies”), 22:21-25 (“The automated bidding application is responsive to information from VOHE 300 describing the video object(s) being auction.”).

82. Conversely, when tracking follows association, the associated information enhances tracking efficiency. Product images linked to objects provide additional visual references—including alternative angles, lighting conditions, and color variations—that improve the system’s ability to track objects consistently across frames. EX1006, 12:4-9 (“Perspective views from such 2D/3D wire frame models may be advantageous to the automatic object recognition process.”). This bidirectional optimization demonstrates why a POSITA would have recognized the value in flexible step ordering. Additionally, a POSITA would have recognized that the value of the hyperlinks at auction would be instructive in determining whether to invest the resources into tracking the item throughout the video. For example, if at the association step, the minimum bid is not met during the auction then the subsequent tracking of the item throughout the video may not be economically justified. Therefore, a POSITA would have recognized the potential savings and efficiency gain from performing the association step’s auction process prior to the tracking step.

83. Furthermore, Ohliger explicitly teaches that “the relative order of steps can frequently be changed without impacting the system” and that Figure 1 “should

not be read as a limitation to a specific order unless it is clear from the context that a particular order is required.” EX1006, 13:6-11. Thus, for example, Ohliger teaches that step 710’s tracking objects across frames could be performed after step 715 and step 720’s association of an object with information.

84. In my opinion, a POSITA would have also considered it obvious to reorder these steps. The technical advantages and any tradeoffs of alternative step sequencing would have been apparent to those skilled in the art. A POSITA would have been motivated to associate an object with database information prior to tracking it across multiple frames because the associated information would facilitate said tracking. This motivation arises from the enhanced recognition capabilities that database information provides.

85. Ohliger’s teachings support this alternative ordering. At step 720, “the video objects are compared with ... 3D (three-dimensional) wire frame data” of that object. EX1006, 12:4-7. Significantly, Ohliger also teaches that “[t]hese 3D wireframe models may be used to improve the recognition of video objects.” *Id.*, 9:53-56. The dual purpose of these wireframe models—both for comparison and recognition enhancement—demonstrates their utility in tracking operations. Accordingly, in my opinion, a POSITA would have been motivated to obtain an object’s 3D wireframe from the object database when the object is first identified

within a frame in order to improve the recognition of that object in subsequent frames.

86. The system provides additional mechanisms for enhanced tracking through early information association. At step 715, video creators can assign pictures to objects, and a POSITA would have been motivated to use these pictures to improve tracking of the object across subsequent frames. The technical rationale is straightforward: supplemental visual data enhances tracking accuracy. In my opinion, a POSITA would have understood that pictures of an object, especially at alternate angles, would improve the system's ability to identify the same object in subsequent frames by providing more data to compare for each object. EX1006, 12:7-9 ("Perspective views from such 2D/3D wire frame models may be advantageous to the automatic object recognition process."), 13:17-21 ("Analysts or video creators may manually train the segmentation application 106 to recognize an object by, for example, tracing the outline of a video object or the system may present patterns of content objects it found by statistical pattern recognition."). A POSITA would have recognized that multiple sources of visual reference data create a robust framework for consistent object tracking throughout the video sequence.

87. In my opinion, a POSITA would have had a reasonable expectation of success, as the reordering of these steps is expressly contemplated in Ohliger. EX1006, 13:6-11. A POSITA would have further understood that the reordering

would have been simple and well within the skill of an ordinary artisan to implement, as it would merely require performing step 715 and 720 as an object is first identified in a frame, and then using such information to track that object in subsequent frames as taught in step 710. As discussed above, a POSITA would have recognized the need for flexibility in associating step timing based on information sources—whether manual human input or automated processes like database queries. In my opinion, these constraints were common for software systems at the time and well within the abilities of a POSITA to address.

**(iv) 1[b][i]: i. identifying one or more of the plurality of visual or audio objects within the video media,**

88. In my opinion, Ohliger discloses this limitation. Ohliger describes a comprehensive segmentation process that produces an object list cataloging all visual elements found within video content—*identifying visual objects within the video media*. The system executes this identification through systematic frame-by-frame analysis. Specifically, Ohliger’s segmentation process 700 segments video and generates “a list of video objects throughout the video (including the temporal coordinates and region definition data for each object).” EX1006, 11:53-60, 16:19-24 (“[T]he video content analysis application 106 segments the raw video content to yield a list of all the video objects in a given video.”). This segmentation yields comprehensive object metadata, capturing both the spatial boundaries of each object

within frames and their temporal presence across the video timeline. The resulting object list serves as the foundation for all subsequent interactive functionality.

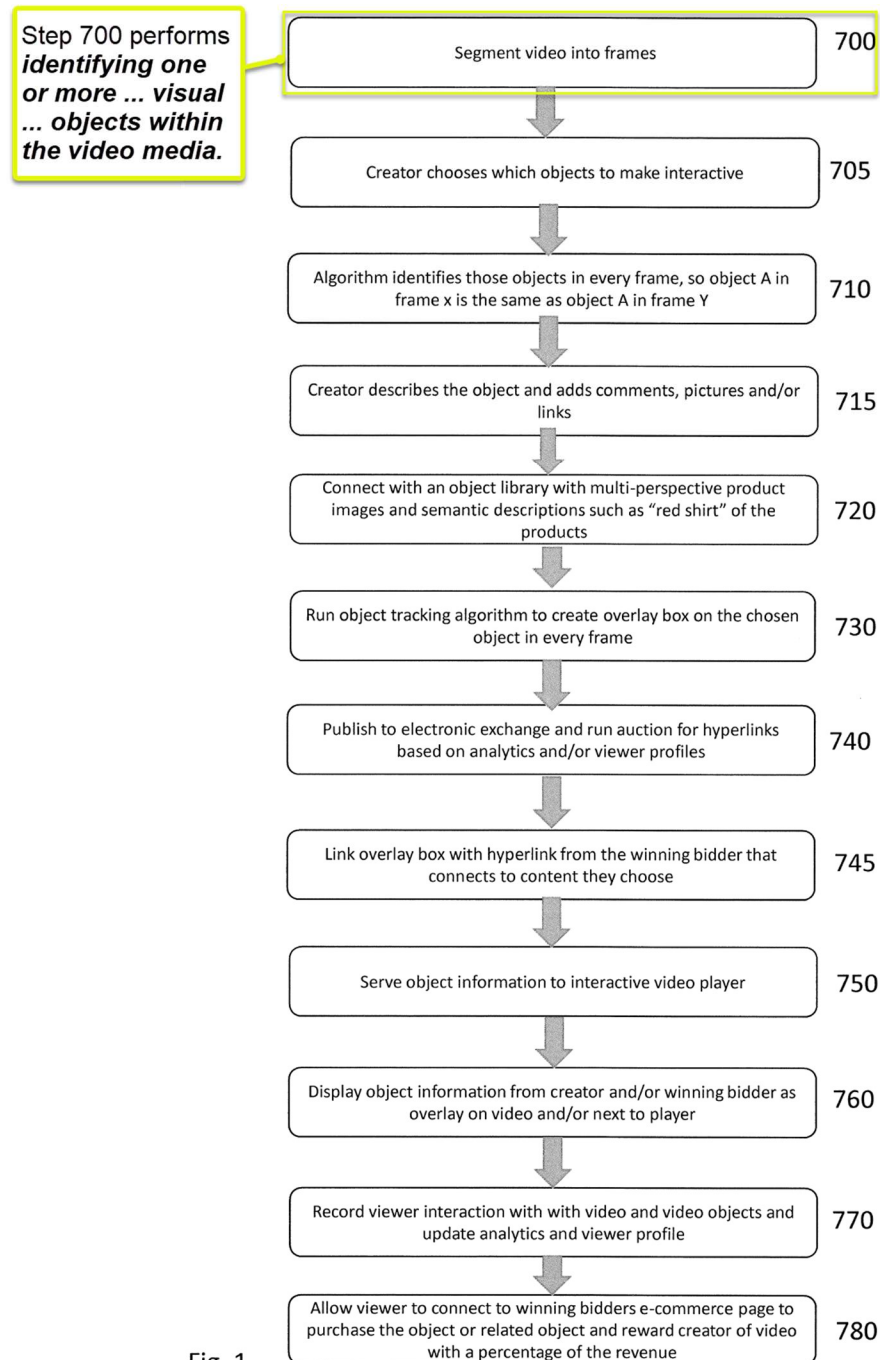


Fig. 1

89. Ohliger discloses multiple technical approaches for object identification within its segmentation application. The system “recognize[s] an

object by, for example, tracing the outline of a video object or the system may present patterns of content objects it found by statistical pattern recognition” or “based on size, shape, color, color density, etc.” using “video content analysis technologies.” EX1006, 13:12-25; 1:56-61 (“Video content segmentation is the systematic decomposition of a motion picture into its individual frames in chronological order, and then deconstructing the frame into its objects (components) such as a person, a shirt, a tree, a leave etc. Segmenting video content results in a large number of objects with little value if not classified.”). In my opinion, a POSITA would have understood that the system described by Ohliger analyzes multiple visual attributes—including dimensional properties, chromatic characteristics, and density distributions—to achieve robust object identification. Further, a POSITA would have recognized that Ohliger’s multi-modal approach to video content analysis ensures accurate object detection across varying video conditions and object types. As described in Ohliger, this segmentation process systematically decomposes the motion picture at both temporal and spatial levels, separating the video into chronological frames, then identifying discrete objects within each frame’s visual field.

90. Ohliger’s “video objects” constitute *visual objects* because they encompass tangible items that appear visually within the video frames. These objects include “data and visualizations of data which relate to content objects of videos

such as a shirt.” EX1006, 2:53-57, 10:3-6 (“The term video object as used in this specification refers to a video frame component, e.g., a shirt, pants, a hat, shoes, a handbag, or a purse which appears in the video or animation.”). The system processes these visual elements as discrete components that maintain distinct visual characteristics within the video. Each identified object represents a visually distinguishable item—whether clothing, accessories, or other consumer goods—that occupies specific regions within the video frames. These *visual objects* form the fundamental units of interaction within Ohliger’s system, transforming passive visual content into selectable, information-rich elements.

91. Additionally, Ohliger also discloses a manual approach for identifying visual objects within the video content. In this implementation, “an analyst may manually identify/configure the object boundaries.” EX1006, 12:9-15; 15:62-64. Ohliger discloses that this manual approach may be performed independent of or in combination with the aforementioned automated, algorithmic techniques for object identification. EX1006, 12:11-17. A POSITA would have recognized that this manual identification process allows human operators to define precise object boundaries when automated recognition may be insufficient or when specific accuracy is required. EX1006, 18:22-35 (“Additionally or alternatively, an analyst 110 may review and refine the boundaries of an unmatched video object, and then

subject the object to a second round of classification and/or second round of matching the object with objects in the object library.”).

- (v) **1[b][ii]: ii. associating product information to the one or more of the plurality of visual or audio objects to correspondingly form one or more items, services, or songs of interest;**

92. In my opinion, Ohliger discloses this limitation. Ohliger describes comprehensive mechanisms for associating product information with objects through three distinct, independent steps: (1) a **Manual Step** where video creators associate descriptions, comments, and related content to video objects; (2) an **Automated Step** where video objects are compared against databases of known items or products and matching item descriptions and related information from the database are associated with video objects; and (3) an **Auction Step** where hyperlinks are auctioned to advertisers to associate their websites with video objects as they appear in videos. Each step individually teaches this limitation and provides alternative pathways for information association. These steps can operate independently or in combination, creating multiple layers of associated data for each object. The associated product information gathered through these various channels is then incorporated into interactive overlays displayed with videos to form items of interest for viewers. This multi-source approach ensures rich, contextual information becomes available for each identified object, transforming simple visual elements into comprehensive interactive shopping opportunities.

93. “*associating product information to the one or more of the plurality of visual ... objects*”:

94. **Manual Step:** First, Ohliger describes manually *associating product information to visual objects* through creator-driven processes in steps 705 and 715 of Fig. 1. In these steps, a video creator exercises direct control over object interactivity by choosing specific objects and associating their own comments or other information to those selected objects. EX1006, 12:17-24 (“The creator of the video can ... choose which objects they want to make interactive” and then “describe, comment and add additional content to an object in the video that they have chosen to make interactive.”) This manual association process enables creators to curate which visual elements become interactive and to provide customized descriptive content for each chosen object.

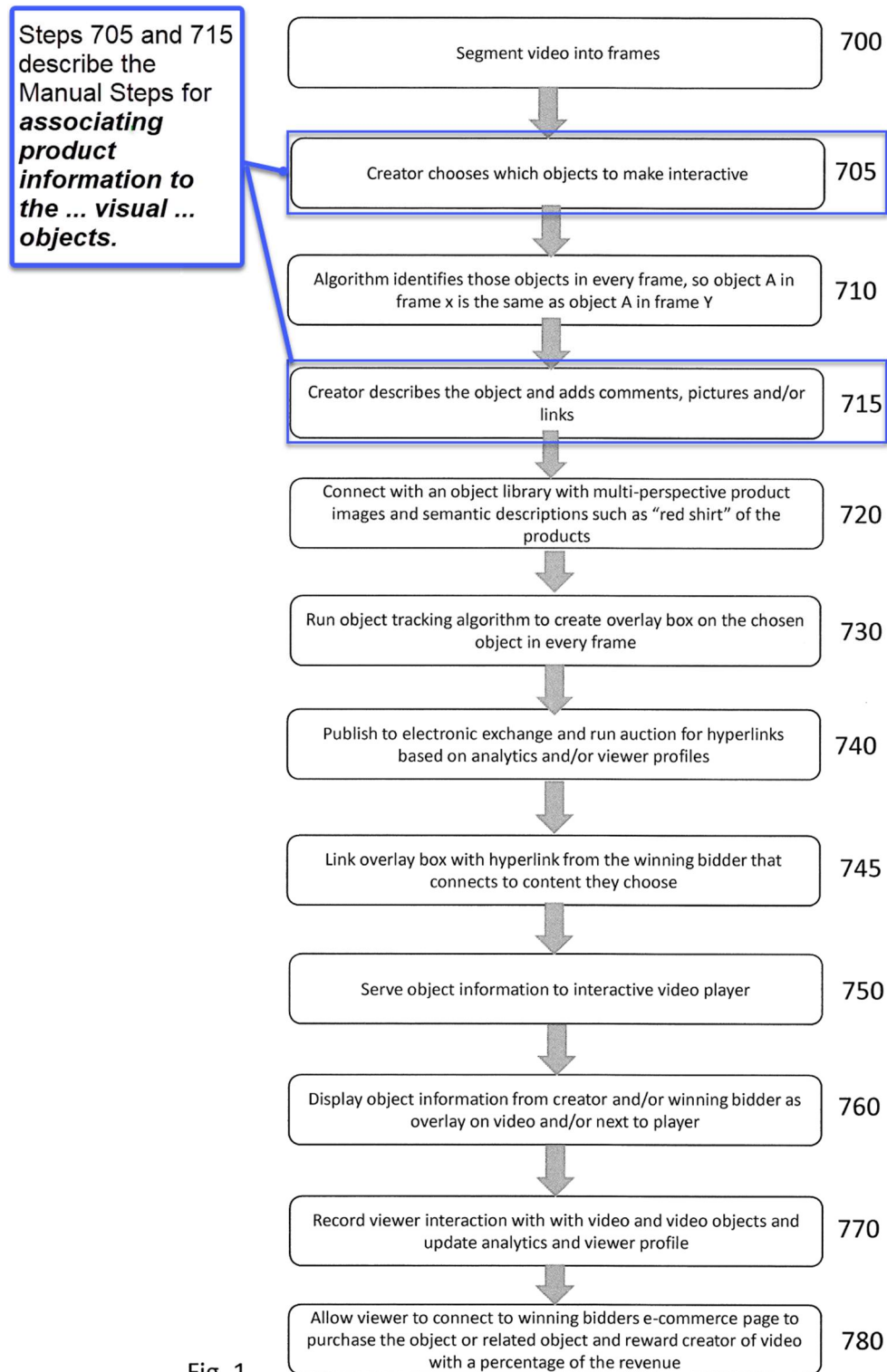
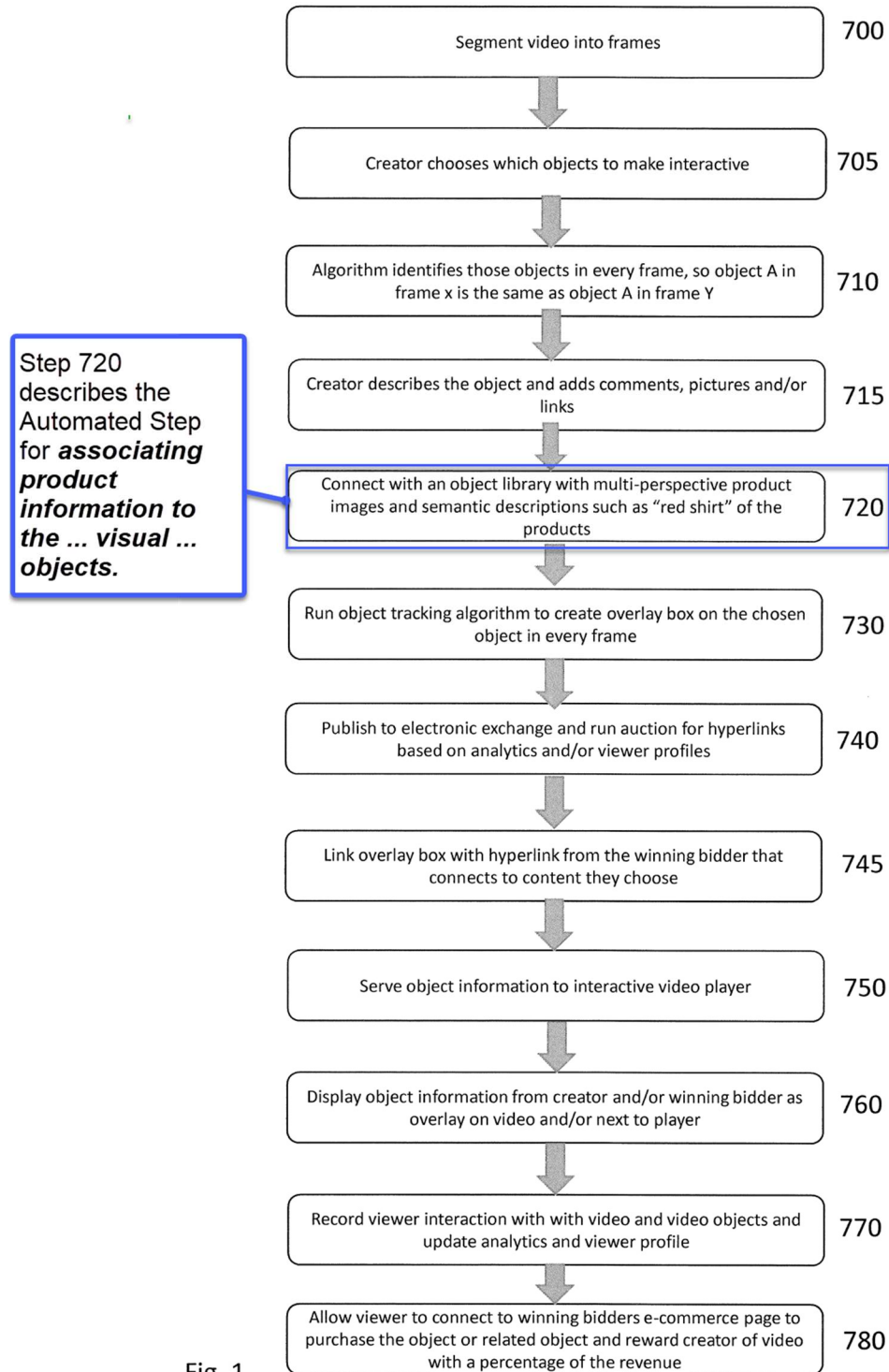


Fig. 1

95. **Automated Step:** In step 720 of Figure 1, Ohliger describes an automated association process where the system compares objects to known objects

recorded in an “object library,” and then associates product information from the “object library” with the video objects identified in the video. This automated matching mechanism compares the identified video objects and entries in the system’s object library based on two-dimensional or three-dimensional wireframe data. EX1006, 12:4-7 (“In step 720, the video objects are compared with objects in an object library, which may be 2D (two-dimensional) or 3D (three-dimensional) wire frame data representing objects within the video.”). When a match is found, Ohliger’s system retrieves the relevant product information from the object library. EX1006, 12:9-11 (“If a match is detected then the object is associated with the product description and/or meta-data from the object library.”), 9:13-15 (“[C]reating a video overlay with region definition data linking relevant content with the video content objects and thereby creating an interactive video.”), 17:27-36. Thus, through this matching process, Ohliger’s system automatically associates both the product description and metadata from the object library to the corresponding video object, establishing the association between visual objects and their product information.



96. **Auction Step:** Ohliger also describes an auction mechanism where retailers, brands, affiliates, or advertisers can associate a website with the visual

object. Through this auction process, commercial entities compete for the right to link their content to identified objects within the video. In step 740 of Figure 1, an “auction allows retailers, brands, and affiliates/advertisers to bid for the right to assign a hyperlink to a video object that takes a viewer of the video who has interacted with that object to a website where that object or any object the winning bidder chooses can be purchased, or to any other content the winning bidder would like to associate with that object.” EX1006, 12:29-37; 10:39-51 (“further aspect of the invention relates to the creation of an electronic auction or exchange for the right to own the hyperlink associated with interactive objects, this exchange will be referred to as a video object hyperlink exchange (VOHE) in this document.”). Therefore, Ohliger’s auction system establishes *associati[ons]* between visual objects and *product information* by enabling winning bidders to attach their

websites and purchasing information to specific video objects.

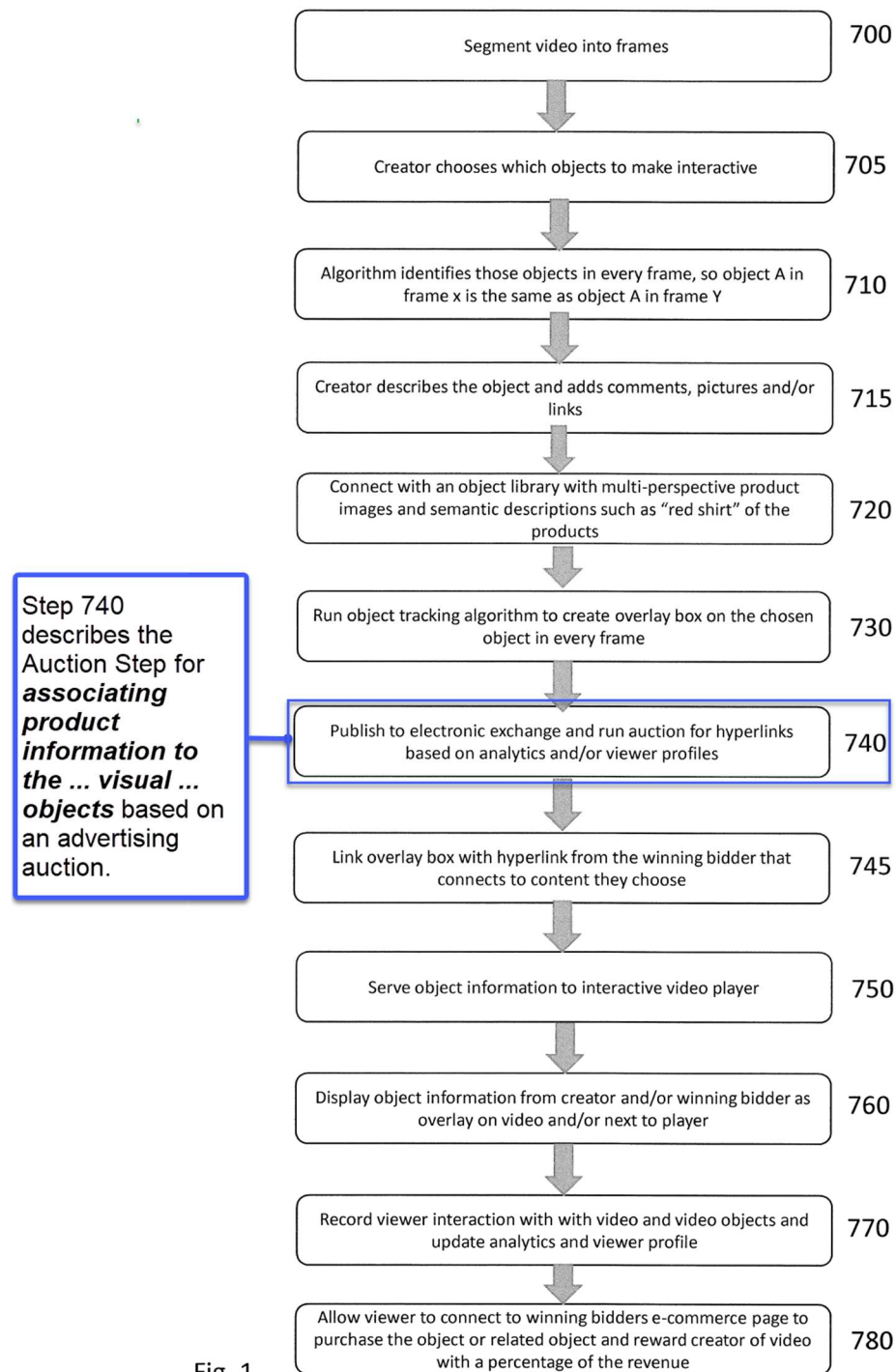


Fig. 1

97. “correspondingly form one or more items, services, or songs of interest”:

98. Through the association of product information with video objects, Ohliger transforms generic video objects into *items of interest* that viewers can learn about and potentially purchase. EX1006, 12:17-28, 12:30-37, 12:42-50, 24:46-57. The transformation from raw visual elements to meaningful interactive content depends on this classification process. Ohliger teaches that objects that are identified are of “little value if not classified.” EX1006, 2:56-61. This limitation exists because an unclassified object could be any object, or no object at all and merely a group of stray pixels.

99. In my opinion, the system in Ohliger would identify a greater number of objects in a video than the creator would want to make interactive because the algorithms for identifying objects would be over-inclusive and produce “duplication/redundancy.” EX1006, 11:54-60. The automated detection processes generate numerous potential objects that require filtering. Additionally, in my opinion, it was well known at the time that the methods of identifying or recognizing objects within video or images, such as the “computer vision algorithms” and “object tracking algorithms” contemplated in Ohliger, may produce false-positives or noisy results. EX1006, 11:32-37.

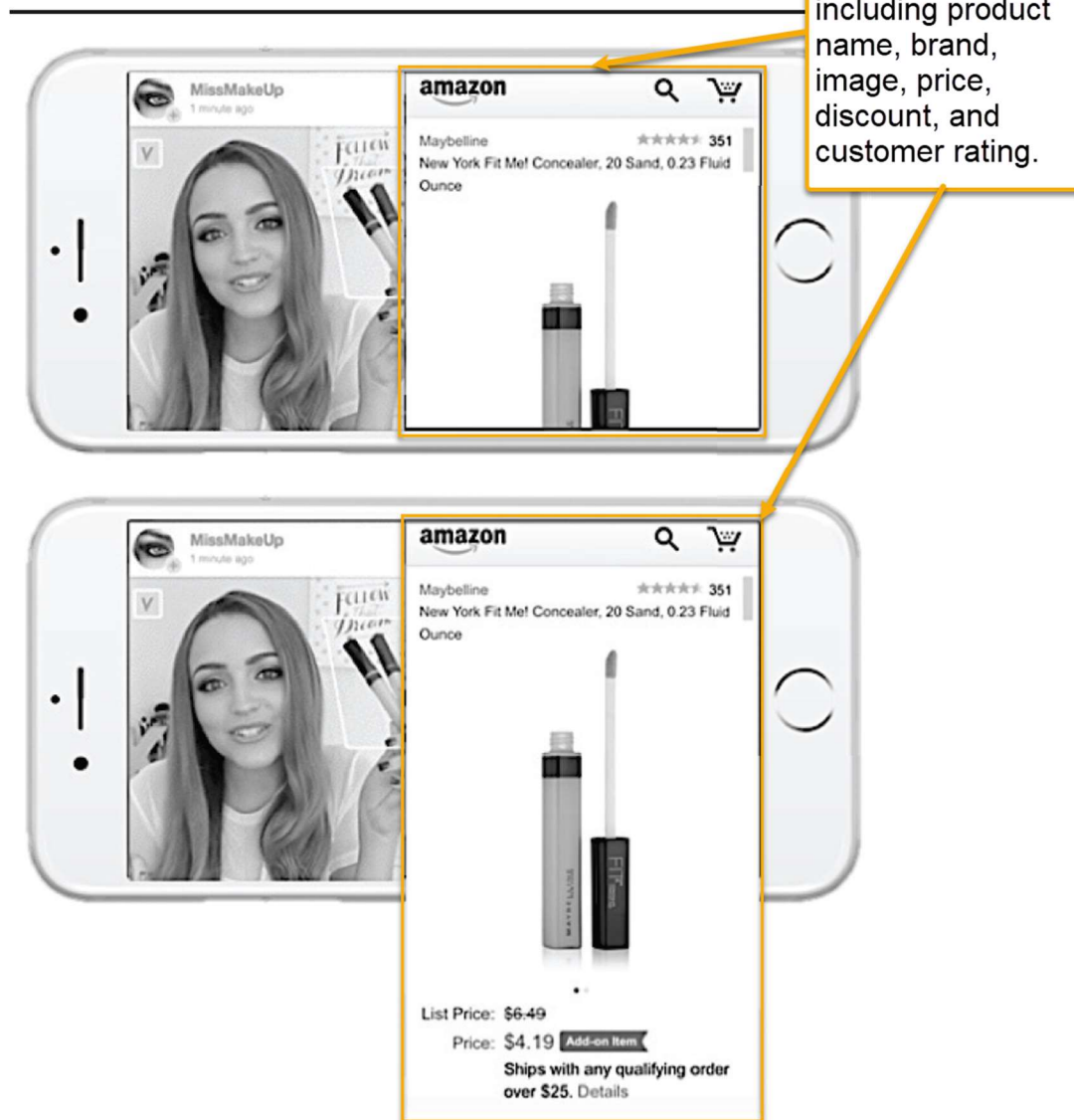
100. However, once the video object is *associat[ed]* with information from the object database, that video object has been recognized as an *item of interest* to viewers. EX1006, 11:38-39 (“[It] is an object of the invention to identify video

objects of interest within video and animations.”). The association process effectively filters the over-inclusive identification results describe above by validating which objects have commercial or informational value. Indeed, Ohliger teaches that “unmatched video object may be discarded,” because video objects without associated information are not *items of interest*. EX1006, 12:11-17. Thus, the *associat[ion]* of *product information* serves as the critical step that distinguishes meaningful *items of interest* from the broader set of detected visual objects. PO advanced the same interpretation during claim construction in the district court case. EX1017, 7 (“[I]t is only when product information is associated with objects that ‘one or more items, services, or songs *of interest*’ are formed.”).

101. Ohliger further teaches that the *associat[ed]* *product information* is displayed upon user interaction. This confirms that the objects have been *form[ed]* as *items ... of interest*. The system presents this associated information when viewers engage with the identified objects. EX1006, 12:24-28 (“[t]his content will appear to a viewer when they interact with the bounding area around the object in any frame that the object appears.”), 14:31-59 (“For example the viewer may point to a video frame, which causes the video to slow down, select or roll-over a video object within the video thereby triggering the VOHE 300 to display contextual information linked or associated with the object as a pop up, overlay or in a field next to the video player.”), 23:63-64 (“[T]he viewer ... encounters video objects of

interest.”), *see also* Limitation 1[b][iv] below. The interactive display of ***product information*** validates that these video objects have been successfully transformed into items of interest because the display after a viewer’s interaction with the video object supports the goal of providing product information relevant to the viewer’s interest. For example, Figure 28 displays an overlay of purchasing information including the product name, price, customer rating, and shipping information for an item of interest. EX1006, 29:19-21, 2:60-65. This comprehensive product information overlay demonstrates how Ohliger’s system presents ***items ... of interest*** with their ***associat[ed] product information*** to facilitate viewer engagement and potential purchase decisions.

Fig. 28  
Buy Via Split Screen



(vi) 1[b][iii]: iii. tracking the one or more items, services, or songs of interest through the one or more frames of the video media; and

102. In my opinion, Ohliger discloses this limitation. Ohliger discloses comprehensive tracking mechanisms for objects across multiple frames within the

video content through at least two distinct implementations: (1) step 710 where a classification process matches objects across multiple frames, and (2) step 730 where an object tracking algorithm creates overlay boxes over identified objects. Each of these tracking steps ensure continuous object identification throughout the video sequence. The first step employs classification-based matching to establish object correspondence between frames, while the second utilizes algorithmic tracking to maintain persistent overlay positioning as objects move through the video. Together, these steps provide robust multi-frame tracking capabilities that enable Ohliger's system to follow objects consistently across their entire appearance duration within the video content.

103. **Step 710:** The classification process analyzes each frame of the video that contains an identified object and compares the object in the frame against other objects found in other frames. EX1006, 11:54-60. Through this frame-by-frame comparison, the system establishes object continuity across the video timeline. Objects matching across multiple frames are classified as the same video object, *e.g.* “that specific red shirt in one frame is the same as the red shirt in another frame.” EX1006, 1:62-65. This classification enables the system to maintain object identity despite movement between frames.

104. Particularly, when objects are first identified in frames, step 710 “automatically detects additional occurrences of the identified object in the video.”

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

EX1006, 13:22-25, 13:32-38 (“For each identified occurrence of an object ... information for identifying the frame (temporal coordinate) and region definition data (location within the frame, e.g. x, y, and z coordinates) in which the object appears ... **to dynamically track an object.**”), 11:21-31, 16:32-44. The tracking system captures both temporal and spatial data for each object occurrence, creating a comprehensive record of object positions throughout the video. Ohliger’s object tracking accounts for “changes in at least one of the perspective and the size and the angle of view in which the corresponding video object is shown in the sequence of frames.” EX1006, 16:45-49; 13:32-38, 18:45-48. Thus, the classification process maintains accurate tracking even as objects undergo visual transformations due to camera movement, object rotation, or distance changes within the video content.

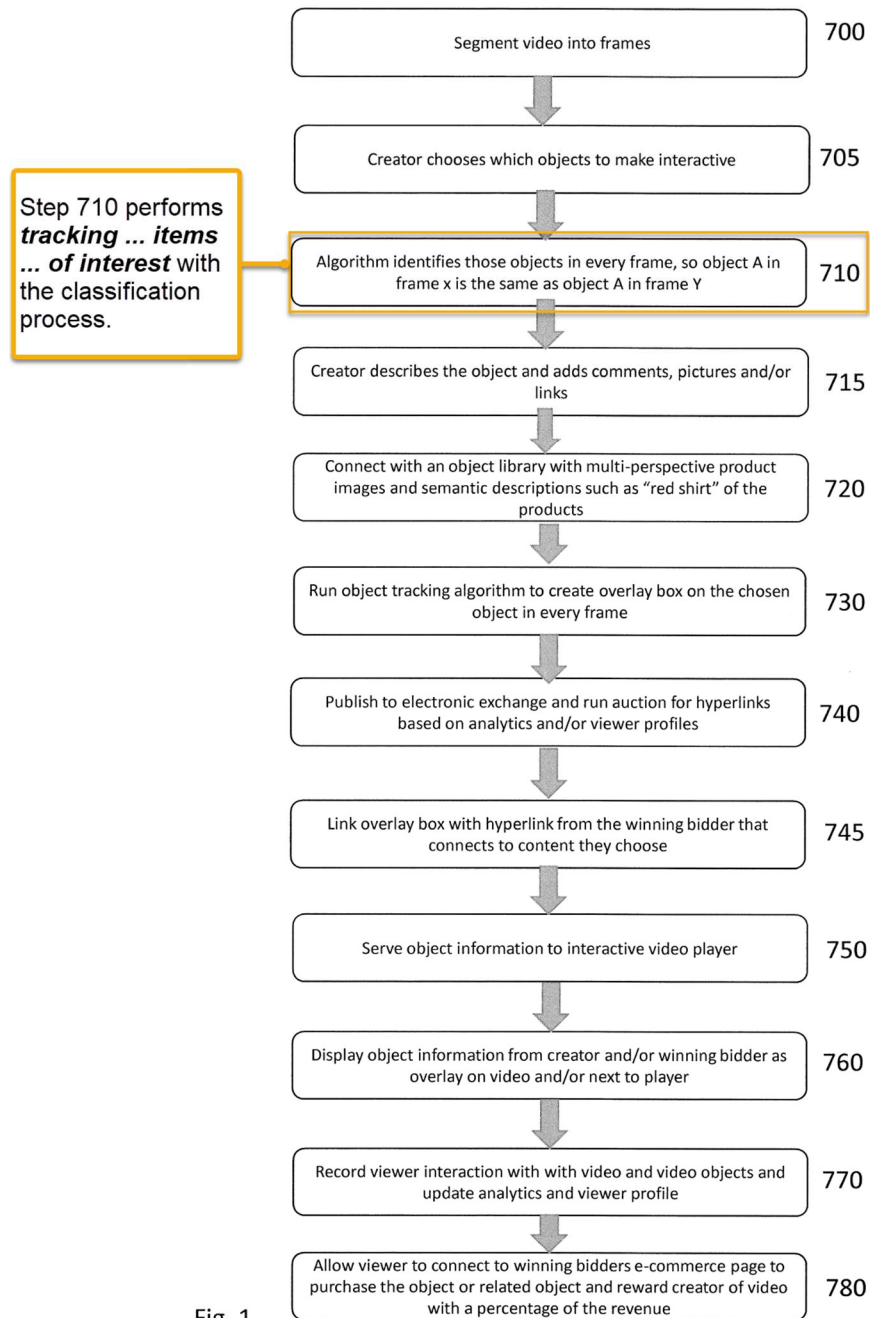


Fig. 1

105. In my opinion, and as discussed above, to the extent Patent Owner argues that this step must be performed after the association of product information, a POSITA would have considered it obvious to perform step 710 after steps 715 and 720. A POSITA would have recognized that the flexibility in step ordering reflects

the technical advantages that different sequences provide. As discussed above, a POSITA would have understood that when tracking follows association, the system can leverage the associated product information to enhance tracking accuracy. The product data, wireframe models, and supplemental images obtained during association provide additional visual references that improve the system's ability to track objects across frames. Therefore, a POSITA would have recognized that reordering these steps to place tracking after association represents a straightforward design choice that optimizes system performance. A POSITA would have had a reasonable expectation of success in re-ordering these steps due to flexibility of the system in terms of constraints and requirement between the steps, and the modular design of the system disclosed throughout Ohliger.

106. **Step 730:** Alternatively, Ohliger discloses an additional “object tracking algorithm” that tracks the video objects in each frame of the video to enable placement of the overlay box which tracks the position of the video object in each frame. This algorithmic approach provides continuous object monitoring throughout the video sequence as part of the overlay generation process. Specifically, Figure 1's step 730 performs an “object **tracking algorithm** to create overlay box on the chosen object in **every frame**.” EX1006, Fig. 1; 12:17-22 (“[T]he object from frame to frame as it is tracked by an object tracking algorithm.”), 32:30-37, Claim 1(f). The tracking algorithm maintains precise positional data for each object as it moves

through successive frames, ensuring that the overlay box remains accurately aligned with the object regardless of its movement or transformation within the video. Through this frame-by-frame tracking mechanism, Ohliger's system preserves the spatial relationship between objects and their associated overlay elements across the entire video duration.

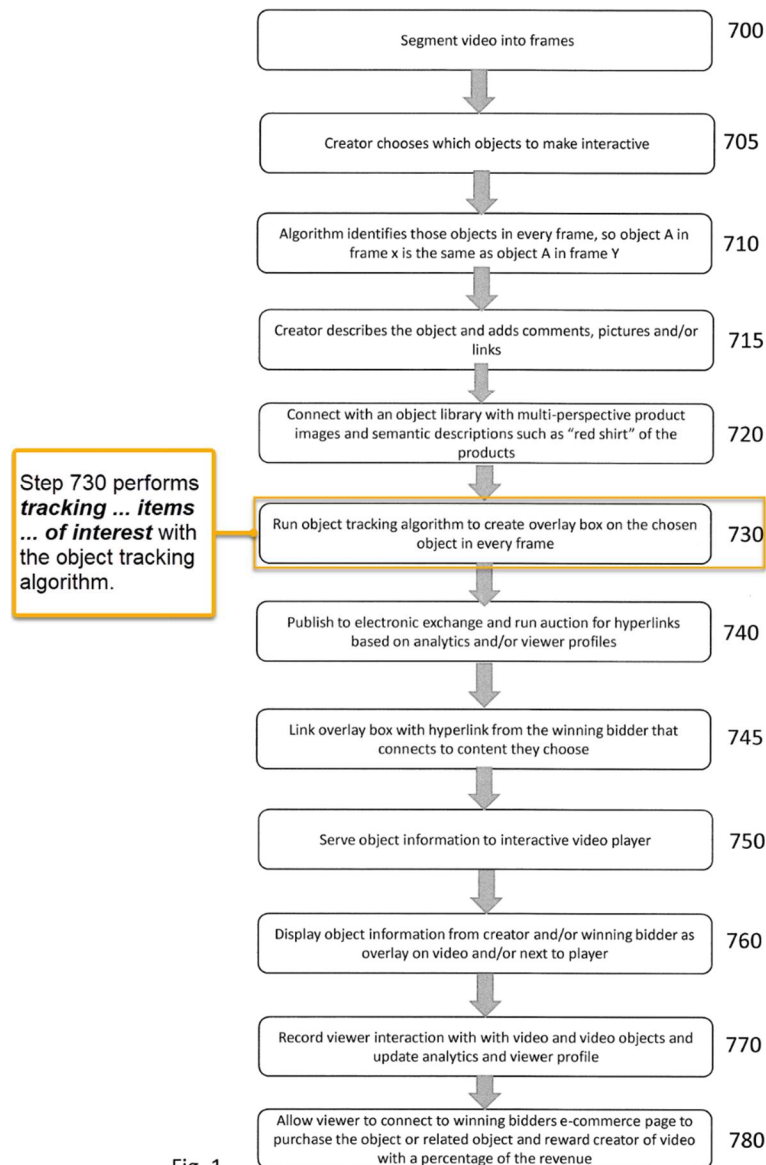


Fig. 1

- (vii) 1[b][iv]: iv. assigning a selectable visual indicator for each of the one or more items, services, or songs of interest, wherein the first interactive video layer is configured to display the selectable visual indicator;

107. In my opinion, Ohliger discloses this limitation. Ohliger describes assigning visual outlines (*selectable visual indicators*) around each recognized video object and displaying these visual outlines in an interactive overlay layer. These outlines define the interactive boundaries for each identified object within the video frames. The outlines are *selectable* and *interactive* because when viewers click, touch, or hover over outlined objects, pop-ups appear over videos to display additional information about selected objects. This interaction mechanism transforms the visual outlines into active interface elements that respond to user input. Through these *selectable visual indicators*, Ohliger's system enables viewers to engage directly with objects of interest while the underlying video continues to play, creating a dynamic interactive experience overlaid on the video content.

108. "*a selectable visual indicator*": Ohliger discloses that the server processes video content by assigning box overlays to identified objects, where overlays are selectable because users can interact with them to show more object information. These overlays serve as the visual interface through which viewers access associated product data. In particular, Ohliger discloses that VOHE server 300 executes an "overlay generation application" that defines overlays tracking

objects over video frames. EX1006, 19:26-29 (“The overlay generation application creates a video overlay including at least one selectable hyperlink whose shape and position (coordinates) track a dynamic location of the video object in the video.”). The dynamic tracking ensures that these selectable overlays maintain alignment with their corresponding objects as the objects move through the video sequence. The overlay can be a box outline, or other shape. EX1006, 16:50-54 (“[T]he outline around the interactive video objects can be a box, a circle, a halo effect, any other shape or effect”). Therefore, Ohliger’s overlay generation application creates *selectable visual indicators* that adapt to various object shapes and maintain their interactive functionality throughout the video playback.

109. The overlay is a *selectable* region, specifically a “clickable space around the object” that causes information to appear when a user “hovers over clicks or touches the object.” EX1006, 33:13-19; 16:63-67 (“[I]n response to a determination that the user action addresses a certain active region, additional information is presented to the user, the additional information pertaining to the video object that corresponds to the certain active region.”), 18:45-48 (“[V]ideo objects are used to create selectable regions (hyperlinks)”). These active regions function as interactive hotspots that detect various forms of user input—whether mouse hover, click, or touch interaction. When the system determines that a user action targets a specific active region, it triggers the display of associated product

information for the corresponding video object. Thus, Ohliger's overlays constitute *selectable* visual elements that transform static video objects into interactive interface components responsive to user engagement.

110. “*assigning ... for each of the ... items*”: The server assigns overlays to each video object, by defining “region definition data” “corresponding” (*i.e. assigned*) to one of the video objects. EX1006, 16:39-44. This correspondence establishes a direct relationship between each overlay and its associated video object. The overlay is then “store[d] ... on the server.” EX1006, 19:29-30, 44-51. The stored overlay data maintains the spatial and temporal information necessary for accurate object tracking. The overlay is assigned to a particular item because it “tracks a dynamic location of the video object.” EX1006, 8:58-9:2, 16:39-44, 33:29-31. This dynamic tracking ensures that each overlay remains bound to its assigned object throughout the video sequence.

111. The assignment of the overlay to an object is further described in step 745, where “a video overlay is created which links the video object with the winning bidders provided content.” EX1006, 12:38-39; Fig. 1 (“Link overlay box with hyperlink from the winning bidder”). Through this linking process, Ohliger's system establishes the connection between the visual overlay element and both the tracked object and its associated commercial content, *i.e.* the hyperlink. Therefore, the overlay assignment creates a persistent relationship that maintains the overlay's

position relative to its assigned object while simultaneously providing the interactive gateway to product information.

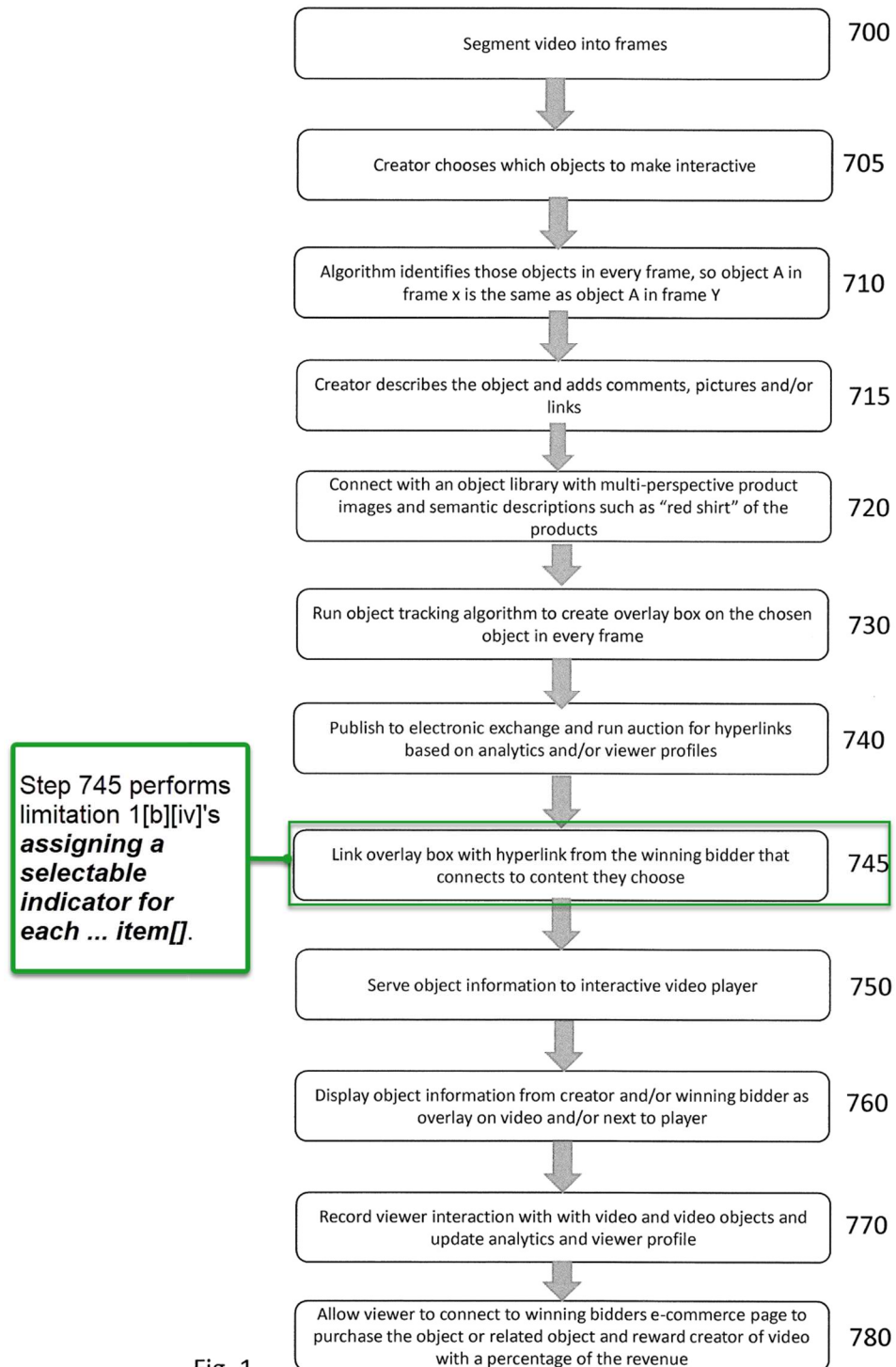


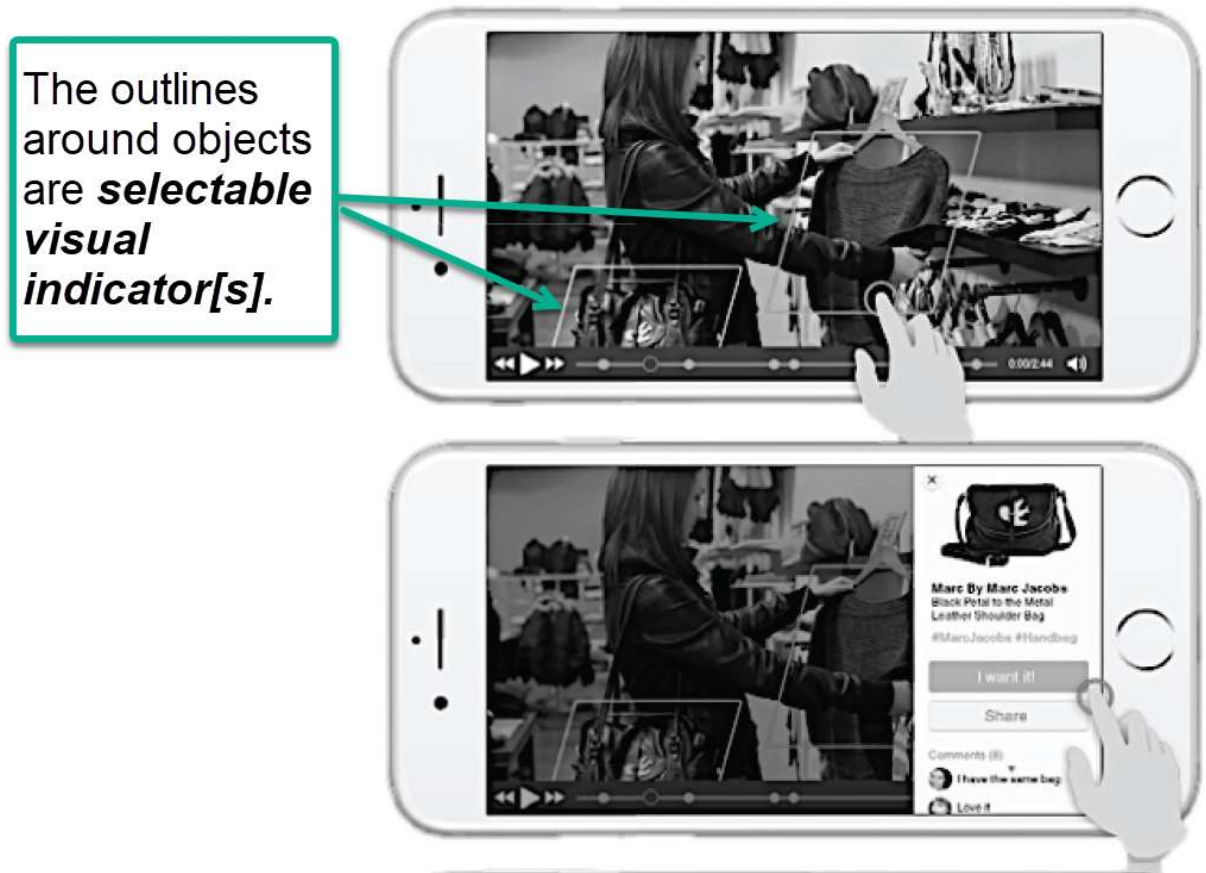
Fig. 1

112. “*wherein the first interactive video layer is configured to display the selectable visual indicator*”: Ohliger discloses that the server then displays the overlay (*i.e. selectable visual indicator*) via an “interactive layer” (*i.e. first interactive video layer*). This interactive layer comprises the selectable overlays within the video content. EX1006, 2:66-3:5 (“[H]yperlinks that are embedded within the interactive layer of video objects and control where a viewer is redirected to if they interact with the video object or objects”), 8:61-63 (“[D]efining region definition data representing each video object on an interactive layer”), 16:32-33 (“[P]roviding active regions for an interactive layer for a video application.”), 16:63-67 (“According to one refinement of the invention, in response to a determination that the user action addresses a certain active region, additional information is presented to the user, the additional information pertaining to the video object that corresponds to the certain active region.”).

113. Figure 18 illustrates the interactive layer displaying *selectable* overlays over video objects. The first screen demonstrates outlines overlaid on visual objects in the video, specifically the parallelograms surrounding the purse and shirt in the video frame. The spatial positioning of the overlays around the objects in the frame indicates that each is associated with their respective overlaid item (*i.e.*, the purse and the shirt). The visual correspondence between overlay locations and object positions confirms the assignment relationship. As shown in the image by the hand

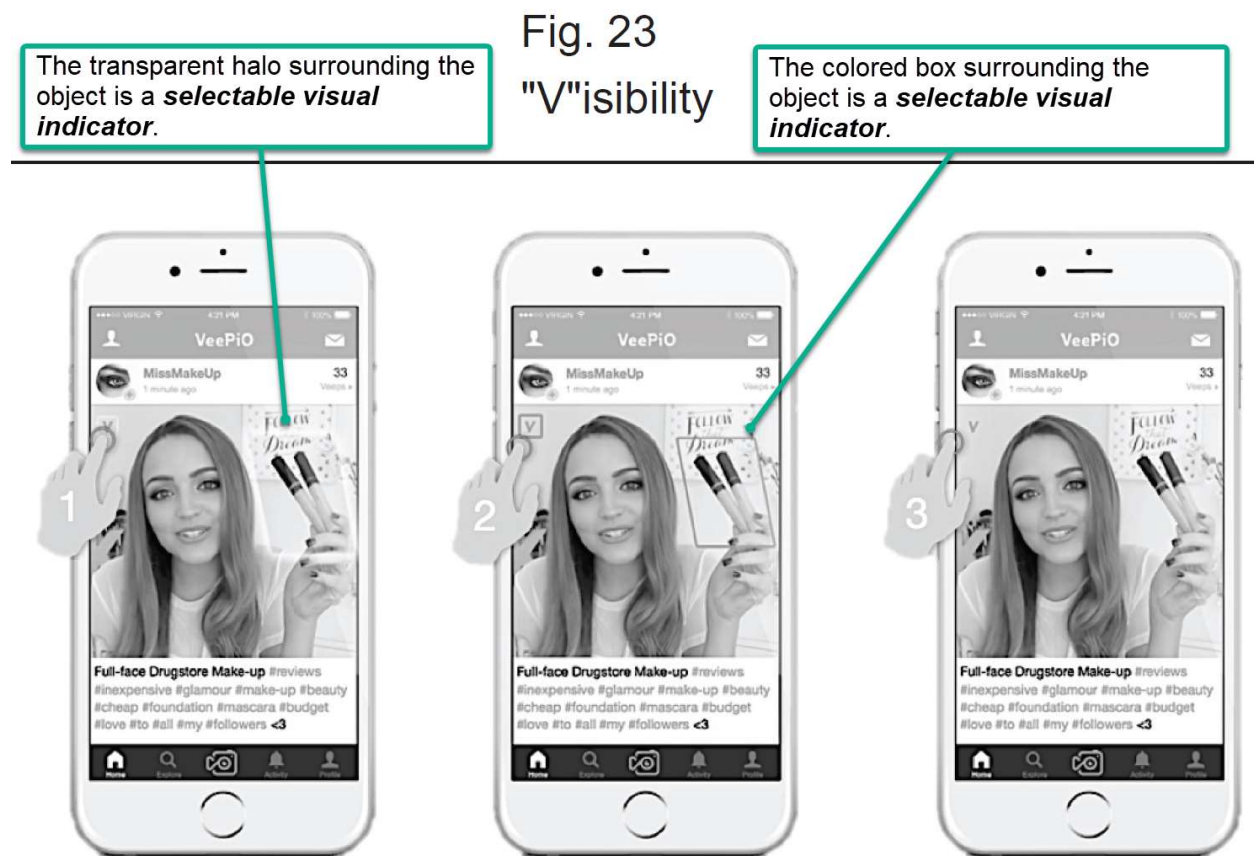
graphic, the user may select the overlay by tapping the object (in this case, a purse).

This user interaction triggers the popup overlay containing the additional information about that object, *i.e.* the purse, in the second screen. EX1006, 28:46-50 (“FIG. 18: VeePiO is simple to use. Tap any Veep’d object and instantly gain the ability to make it your own. (1st screen) Veepio users can also share and view comments for individual objects. (2nd screen) Tapping ‘I want it!’ walks users through an easy in-app purchase flow. (2nd and 3rd screen)”). Therefore, Figure 18 demonstrates how Ohliger’s interactive layer presents *selectable* visual indicators that enable direct user engagement with identified objects to access associated product information.



114. Figure 23 presents additional examples of *selectable* overlays, including “a transparent halo” (1st screen) and “colored boxes on each object” (2nd screen). EX1006, 29:5-10 (“FIG. 23: Once published, the video shows up first in the feed. Tapping the ‘V’ once reveals a transparent halo around each Veep’d object. (1st Screen) Tapping twice places colored boxes on each object so users can view multiple objects with ease. (2nd Screen) Veep’ers can turn objects off by tapping the ‘V’ watermark a third time. (3rd Screen)”). 16:50-53 (“According to one refinement of the invention the outline around the interactive video objects can be a box, a circle, a halo effect, any other shape or effect or no affect at all.”). These varying overlay

styles demonstrate the system's flexibility in visual indicator presentation. The implementation of different overlay formats—from subtle halos to prominent colored boxes—enables the system to accommodate various display preferences and video contexts. EX1006, 16:50-54. Through these alternative visual representations, Ohliger establishes that *selectable* overlays can assume multiple forms while maintaining their core interactive functionality.



(viii) 1[c]: c) generating a video data comprising the video media, the first interactive video layer, and a second interactive video layer,

115. In my opinion, Ohliger discloses this limitation. Ohliger describes comprehensive mechanisms for *generating ... video data* that includes the video

content itself (*video media*), the video objects box overlays (*first interactive video layer*), and the social media overlays (*second interactive video layer*). Ohliger's system architecture supports *video data* generation through disclosing multiple alternative server configurations. Ohliger describes that the **video consumer server**, **broadcaster server**, and/or **video server** can generate this *video data*. The generation of video data occurs after video processing, as the processing creates the overlays that become integrated components of the video data, as discussed above regarding limitation 1[b][i] through 1[b][iv].

116. *First*, Ohliger discloses a “video consumer server” which can generate video data by rendering the overlay onto the video content. EX1006, 19:30-37 (“[T]he overlay is supplied to the video consumer server and is overlaid or superimposed over the video presentation.”), 20:15-20. This rendering process combines the visual elements into a unified video presentation.

117. *Second*, Ohliger describes that the broadcaster server can provide the video content “along with” the overlays. EX1006, 22:7-11. In my opinion, a POSITA would understand that the broadcaster server therefore generat[es] video data that together includes the video and the overlays. The system enables distribution through multiple channels. Ohliger discloses broadcasters as distributing the interactive videos “via interactive TV or streaming websites.” EX1006, 15:41-44.

118. A POSITA would have recognized that these broadcaster channels would have particular requirements for compatible content for “interactive TV” or “streaming.” The technical implementation requires specific format considerations. A POSITA would have recognized that enabling Ohliger’s interactive overlays on the viewer’s device requires software support for the interactive features. As was well-known at the time, digital multimedia formats served as containers that encapsulated the video and audio content as well as supplemental content such as subtitles, navigation menus, images, hyperlinks, and interactive advertisements. EX1021, [0004], [0020]. Apple’s QuickTime video files, for example, included “sprite tracks” for providing interactivity in video content. EX1020, [0242]-[0245].

119. A POSITA would have understood that the broadcaster would have dictated the multimedia format compatible with their broadcasting system, e.g. interactive TV or a streaming website. Therefore, a POSITA would have understood that broadcasters in Ohliger’s system would generate the combined video data through incorporating the overlays into these multimedia formats because it would have integrated into the broadcaster’s existing systems that already had compatible viewer applications and widespread support. EX1021, [0023], EX1020, [0005] (“Quicktime or Adobe Flash media players ... have a large installed base and are generally provided free of charge”).

120. In addition, even if the video and the overlays are separately sent as different pieces of data, this teaching would still meet the claims under Patent Owner's construction of the term. Specifically, PO contends that this limitation is met regardless of how video media and the two layers are provided to a user device (which includes transmitting them separately), as long as they are displayed together. EX1005, 32-35 (alleging limitation is met based on the display of video and overlays at a user device without any analysis or allegation addressing how they are transmitted). Ohliger discloses that the video content and overlays are displayed together, concurrently with the overlays appearing over top portions of the video content. EX1006, Fig. 18.

121. *Third*, Ohliger describes that the video server can also provide the video content "along with" the overlays. Ohliger specifically describes a configuration that includes "video server 200,300" which performs the function of both the VOHE server 300 and the broadcaster server 200. EX1006, 26:31-27:26 ("The system includes (and executes) a software application for accessing and displaying the at least one video file"). This unified server architecture consolidates the video processing and distribution functions into a single system. In this configuration, the video server performs the function of generating video data that includes the video content "along with" the overlays. EX1006, 22:7-11. Therefore, the video server

200,300 combines both the video content and overlay generation capabilities, producing integrated video data that contains all interactive elements.

122. Alternatively, it would have been obvious for the VOHE server 300 to generate the video data by rendering the overlays onto the video content. The technical feasibility of this approach stems from the server's existing architecture. Ohliger discloses that the VOHE server 300 stores both the video content and the overlay layers. EX1006, 26:33-46, 15:59-67, Figure 2. A POSITA would have recognized that with these components already resident on the server, combining them represents a straightforward technical implementation. The server-side generation of integrated video data would leverage the VOHE server's existing storage and processing capabilities.

123. A POSITA would have been motivated to perform server-side rendering prior to transmission to the viewer to prevent well-known techniques that block advertising overlays based on network origin. EX1010, 11:55-59. A POSITA would have understood that this technical approach protects the advertising capabilities, a significant commercial concern in Ohliger's system. The economic incentives for server-side rendering align with Ohliger's disclosed business model for the system. Indeed, Ohliger is directed to auctioning off object overlays to advertisers, and preventing the overlays from being blocked would provide the benefit of increased revenue to advertisers and therefore higher auction prices for

video creators. This protection mechanism ensures the integrity of the monetization system that underlies Ohliger's interactive video platform.

124. In my opinion, a POSITA would have been motivated to perform centralized rendering at the video server rather than the user device as the video server has more powerful rendering hardware and the render need only be performed once instead of being rendered by each individual user device, thereby increasing efficiency and speed. A POSITA would have recognized that the computational advantages of server-side processing. Additionally, a POSITA would have known that the centralized approach eliminates redundant processing across multiple client devices. As discussed above, a POSITA would also have been motivated to integrate the overlays into existing multimedia formats for broadening support for the interactive functionality. EX1021, [0004], [0020], [0023]. Therefore, a POSITA would have recognized that server-side rendering at the VOHE server 300 represents an obvious implementation choice that leverages existing capabilities while providing technical and commercial benefits.

125. Moreover, in my opinion, a POSITA would have been motivated to provide the video content and the interactive video layers from VOHE server 300 to viewer 124. This direct distribution architecture offers multiple technical and commercial advantages. Since VOHE server 300 already stores all of those components, having VOHE server 300 provide them to the viewer 124 would

obviate the need for VOHE server 300 to first send the interactive layers to broadcaster 120. A POSITA would have recognized that eliminating this intermediate transmission step reduces latency and network overhead.

126. The direct connection between VOHE server 300 and viewers enables enhanced data collection capabilities. A POSITA would have recognized that directly providing the video content and interactive video layers to the viewer would allow for building viewer profiles with a larger dataset on each viewer as opposed to relying on the broadcaster to share viewer profiles. EX1006, 14:4-21, 14:60-67, 22:61-23:3. The system's profiling mechanisms leverage multiple data sources. Ohliger discloses creating viewer profiles based on data collected through "cookies stored on the viewer's computer, or from third party information of the viewer." EX1006, 14:12-21.

127. A POSITA would have been motivated to build more robust viewer profiles to improve the ability of advertisers to target their audience for the hyperlink auction. EX1006, 14:18-21, 14:60-67. The enhanced targeting capabilities would increase the commercial value of the advertising placements. A POSITA would have recognized that having the system of Ohliger directly connect with the viewer's device, as opposed to using the broadcaster as a middleman, would have allowed for gathering information on the user's interactions with the video. This direct

connection provides granular interaction data that would otherwise be unavailable through intermediary distribution.

128. In addition, where the video content originates from a content creator who is just another user of the Veepio application, the creator likely does not have the same uptime as VOHE server 300 (*e.g.*, the creator may not always have their phone powered on or with Internet connectivity); thus, having VOHE server 300 be the source of the video content, in addition to the interactive layers, to other viewers would improve availability of the video content. A POSITA would have understood that centralized hosting addresses reliability concerns inherent in peer-to-peer content distribution. Indeed, this was the framework already employed in social networking sites discussed in Ohliger including Instagram, Vine, Twitter, and Snapchat. EX1006, 25:47-54, 15:36-44. These platforms demonstrate the established industry practice of centralized content hosting for user-generated video content.

129. In my opinion, a POSITA would have had a reasonable expectation of success. The technical implementation involves established processes familiar to those skilled in the art. As discussed above, since the requisite components (video content and overlays) are already resident on the VOHE server, performing the rendering on the server would entail a routine conversion of information in one format to another, which was well-within the skill of an ordinary artisan. EX1021,

[0023], claim 3. A POSITA would have recognized that combining video streams with overlay data represents a standard multimedia processing operation. The server-side rendering techniques required for integrating overlays with video content were mature technologies at the time of Ohliger. Therefore, in my opinion, implementing the proposed modifications would have presented no technical barriers that would discourage a skilled artisan from pursuing this approach.

**(ix) 1[c][i]: wherein the video data is configured to be provided to a user via a first media playback device and to at least one guest via a second media playback device,**

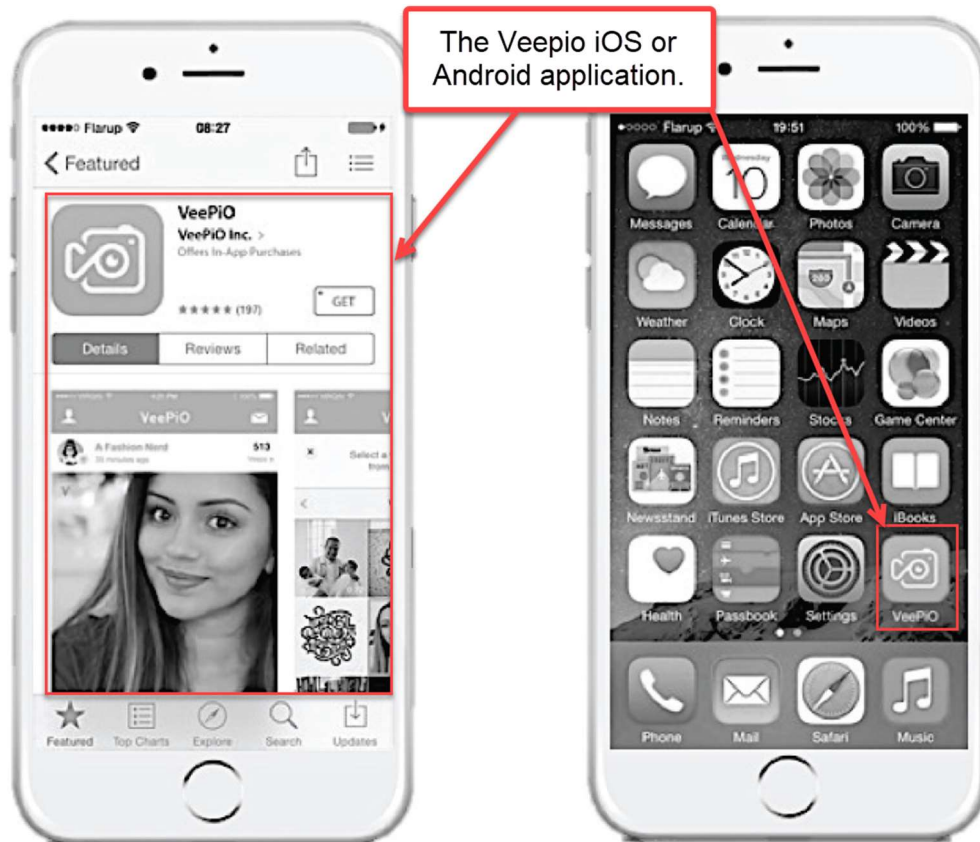
130. In my opinion, Ohliger discloses this limitation. Ohliger discloses comprehensive distribution mechanisms where video content with overlays (i.e. *video data*) is provided to multiple users via their respective media playback devices such as through an iOS or Android application. The system supports simultaneous multi-user access to the interactive video content. Ohliger discloses that each video has multiple “viewers” which can interact with the overlays and communicate with each other. EX1006, 9:5-11 (“Allowing viewers of the of the video to interact with objects in the video by clicking or touching areas around objects in the video”), 11:11-14 (“The invention also includes giving video viewers the ability to comment and communicate with each other”), 30:55-60.

131. The social sharing capabilities enable viral distribution of the interactive content. Ohliger also discloses that a first viewer (user) has “the ability

to share the video” with a second viewer, who may be a “friend” on social media (guest). EX1006, 24:50-57, 28:35-37. This sharing mechanism facilitates the expansion of viewership beyond the initial user base. Each user views the video via “a computer, mobile phone or another electronic device.” EX1006, 22:28-33. Therefore, Ohliger’s system provides video data to multiple media playback devices operated by different users, including both primary users and their invited guests, creating a distributed viewing environment for the interactive video content.

132. Alternatively, Ohliger discloses that video creators or broadcasters (*the user*) can provide videos and overlays to viewers (*guest*). EX1006, 15:36-44 (“[B]roadcaster’ simply refers to the party who owns the rights to the video content and makes it available to viewers 124”), 22:7-11, 34:16-21. This distribution model establishes creators and broadcasters as primary content providers who control access to the interactive video experiences. However, both creators/broadcasters and viewers use the same “VEEPIO app” to view and interact with videos. EX1006, 34:41-44; Fig. 13. A POSITA would have recognized that this unified application architecture enables seamless transitions between content creation and consumption roles. The shared platform ensures that all participants—whether distributing content as users or receiving it as guests—access the same interactive features through identical media playback devices. Thus, Ohliger’s system supports bidirectional

content flow where creators distribute video data to multiple viewer devices while maintaining consistent interactive functionality across all user types.



133. After or as part of uploading videos, creators (*users*) can access videos through the same user interface and leave comments like any other viewers (*guest*). EX1006, 11:11-14, 30:55-64; Fig. 22, 29:1-4. This dual functionality demonstrates that creators operate both as content providers and consumers within the platform. The system architecture enables creators to engage with their own uploaded content using the identical viewing and commenting features available to all platform participants. Thus, creators (*user*) are provided video content and associated interactive layers (*video data*) like any other viewers. A POSITA would have

recognized that this symmetric access model ensures creators experience the same interactive capabilities they provide to their audience.

134. These disclosures in Ohliger are consistent with Patent Owner's infringement contentions, which allege that "users of TikTok iOS and Android apps" are both "*the user*" and "*at least one guest*." EX1005, 35-40.

- (x) 1[c][ii]: wherein the second interactive video layer comprises a social media portion controllable by the user via the first media playback device, so as to interact with the at least one guest that is also a viewer of the video media through an interface displayed in the second interactive video layer via the second media playback device,

135. In my opinion, Ohliger discloses this limitation. Ohliger discloses an *interactive layer* where users (*i.e. the user and guest*) interact with one another by using social media to comment about objects found during playback of the same video with the same overlays to users via their respective media playback devices. This social interaction functionality transforms the video viewing experience into a collaborative environment. Video objects each have their own thread of comments, which Ohliger terms these threads "sub-comments," where viewers contribute to the conversation for a particular object in the video. A POSITA would have recognized that this object-specific commenting architecture enables focused discussions about individual items within the video content. The granular commenting system allows viewers to engage in targeted conversations about specific products or items of

interest. Each viewer may view the comments for each object by selecting the outlined overlay around the particular object when it appears in the video. Therefore, Ohliger's *interactive layer* facilitates multi-user engagement through object-specific social interactions, enabling both *the user and guest* to participate in shared discussions while viewing the same interactive video content on their respective devices.

136. *“wherein the second interactive video layer comprises a social media portion”*: Ohliger discloses a second interactive layer where users can communicate with one another about objects that were selected within a video via the object's box overlay. This communication layer operates independently from the first interactive layer's shopping functionality. EX1006, 11:11-19 (“The invention also includes giving video viewers the ability to comment and communicate with each other and with the creator of the video in regards to a specific interactive object in the video. This is referred to as ‘sub-comments’ and it differs from current comments on the internet and on social media in that sub-comments are accessed by viewers within the content related to the interaction of the viewer with a specific object within a video as opposed to comments that relate to the video as a whole.”).

137. The sub-comment architecture provides object-specific social interaction. A POSITA would have recognized that this system enables granular, contextualized discussions tied directly to individual visual elements within the

video stream. Unlike conventional video commenting systems that apply globally to entire videos, Ohliger's sub-comments create discrete conversation threads for each identified object. These object-specific threads maintain their association with the visual objects as they appear throughout multiple frames. Therefore, the second interactive layer provides a social communication framework where *the user and guest* can engage in targeted discussions about specific items while those items remain visible and selectable through their overlay indicators.

138. Ohliger describes these interactions as a "community," thereby making them a *social media portion*. EX1006, 28:42-45 ("VeePiO thrives on community. Commenting is easy and encouraged via standard pathways like the feed, and through product sub-comments. Beyond that VeePiO media can be shared anywhere online."). The community-driven architecture transforms individual viewing experiences into collective social engagement around shared content. A POSITA would have recognized that this community framework establishes the social infrastructure necessary for multi-user interaction within the video platform.

139. Indeed, Ohliger's system integrates with well-known social media platforms. EX1006, 25:47-53 (describing integration with Facebook, Twitter, Instagram, Vine, and Snapchat). These integrations demonstrate the system's compatibility with existing social media ecosystems. A POSITA would have understood that leveraging established social platforms extends the reach of object-

specific discussions. Through these social media connections, Ohliger enables sharing and discussion of interactive video content, confirming that the second interactive layer functions as a *social media portion* where *the user and guest* can engage in community-based interactions around identified objects.

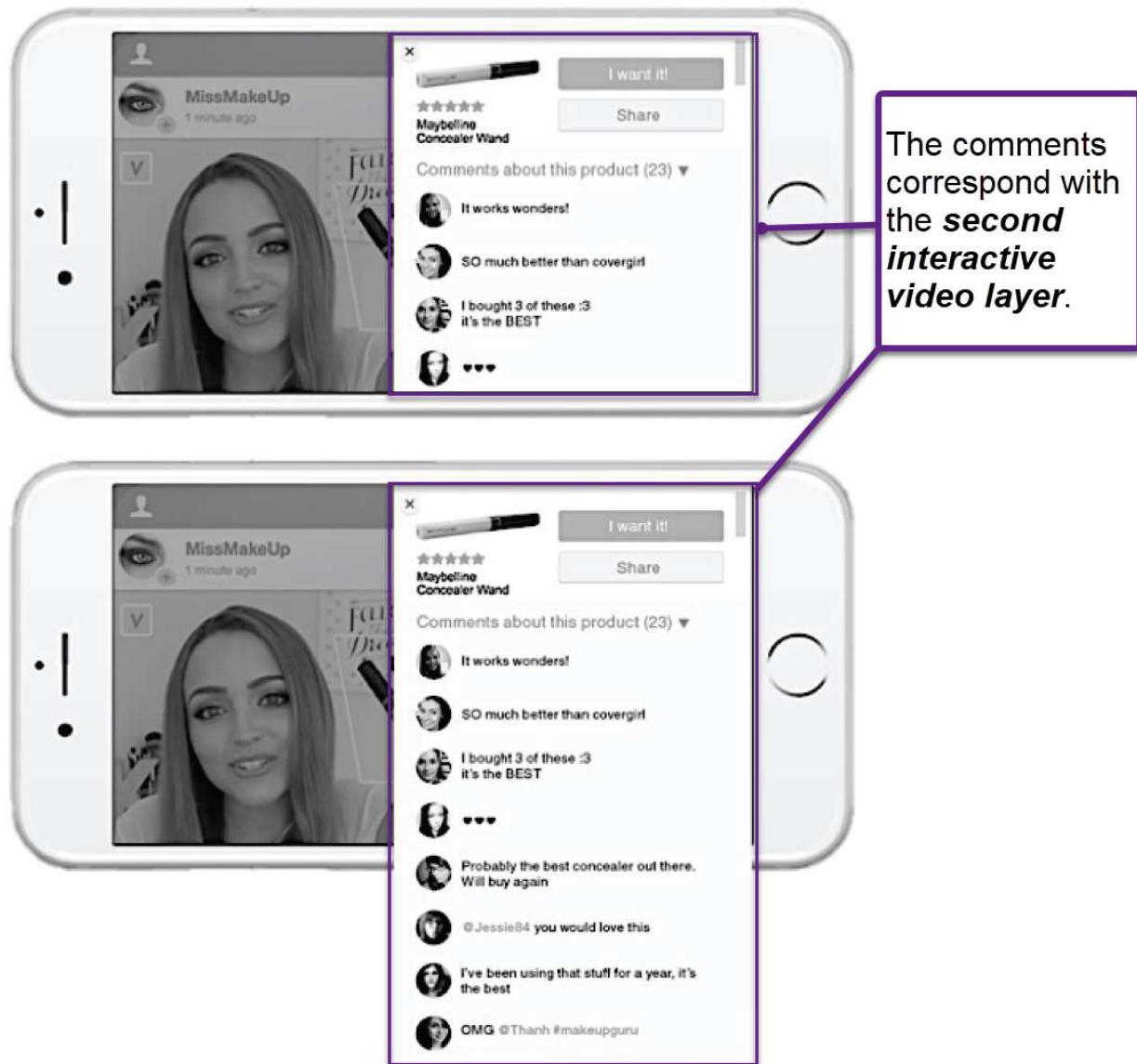
140. Consistent with these disclosures, both the '316 patent's specification and Patent Owner's Infringement Contentions describe user comments as conversations within a social media portion. EX1001, 3:43-46 ("The second of the interactive video layers will provide a public area for users to see a public feed of viewers' comments and reactions to the video media."); EX1005, 40-46 ("a second interactive video layer consisting of a comment section."). The Patent Owner's identification of the comment section as the *second interactive video layer*, aligns with Ohliger's object-specific commenting functionality.

141. Ohliger teaches that these comments are displayed on an interactive overlay layer in the video. The visual presentation of comments directly within the video viewing area creates an integrated *social media* experience. For example, Figure 26 shows the users' comments as an overlay to the video. EX1006, 29:15-16; 24:46-57 ("Once the object has been selected ..., the viewer displays an abstract of the video object in the form of an overlay on top of the current viewing window [which] include[s] ... comments by other viewers"). A POSITA would have recognized that this overlay presentation technique ensures comments remain visible

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

and contextually connected to the specific objects being discussed. Through this overlay mechanism, Ohliger demonstrates that the *second interactive video layer* containing the *social media portion* appears concurrently with the video content, enabling real-time social engagement while users view the interactive objects.



142. Ohliger also “[a]llow[s] the creator of the video to associate comments ... with interactive areas around objects in their videos” and “allow[s] viewers to

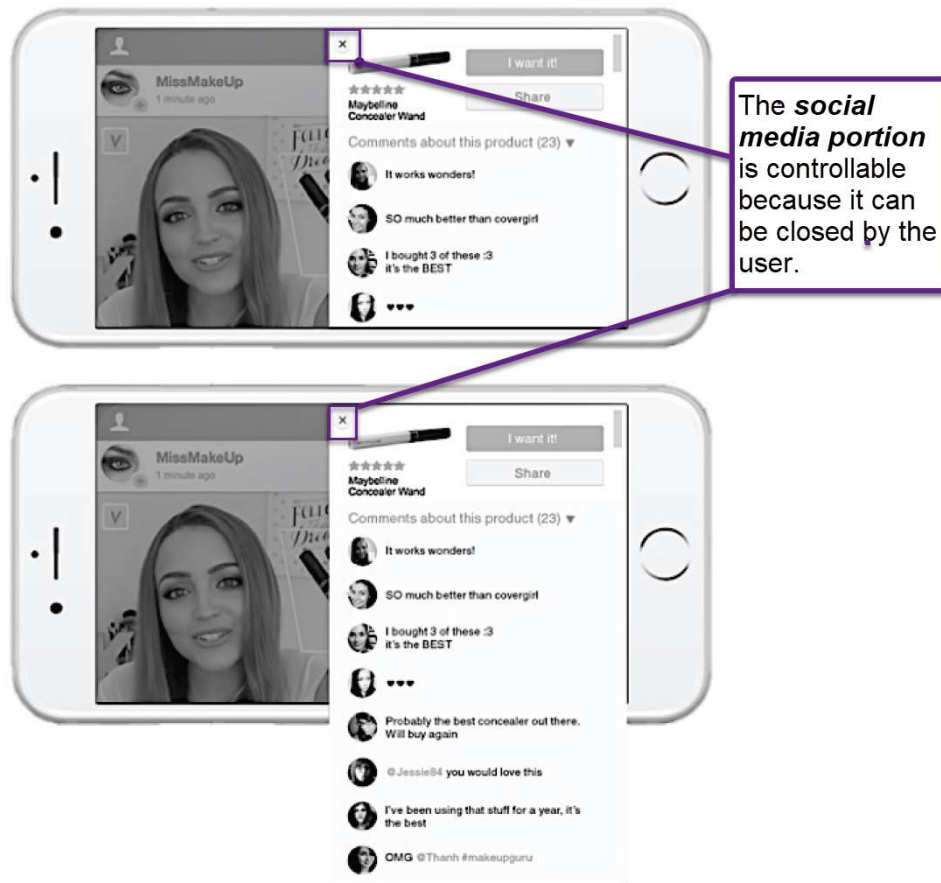
comment about those objects.” EX1006, 9:2-11. This bidirectional commenting capability enables both content creators and viewers to contribute to object-specific discussions. The system architecture supports multiple participants adding their perspectives to the same object threads. For example, Figure 18 explains: “FIG. 18: VeePiO is simple to use. Tap any Veep’d object and instantly gain the ability to make it your own. (1st screen) **Veepio users can also share and view comments for individual objects.** (2nd screen) Tapping ‘I want it!’ walks users through an easy in-app purchase flow. (2nd and 3rd screen)”. EX1006, 28:46-50. A POSITA would have recognized that this implementation creates a commenting system where all platform participants—creators and viewers alike—contribute to the same object-specific conversation threads. The ability to both share and view comments for individual objects establishes the interactive nature of the *social media portion*. Through these mechanisms, Ohliger demonstrates that the second interactive layer supports multi-directional social engagement, where *the user* and *guest* can both post and read comments associated with specific visual objects, creating a dynamic social environment overlaid on the video content.

143. “*controllable by the user via the first media playback device*”: Ohliger discloses that the social media portion is *controllable by the user* because it is activated when the user interacts with the object’s box overlay. EX1006, 11:14-19 (“[S]ub-comments are accessed by viewers within the content related to the

interaction of the viewer with a specific object”), 22:42-48 (“Interaction by the video viewer 124 triggers display of ... comments by other viewers”). This user-initiated activation mechanism places control of the social media functionality directly in the hands of each viewer. A POSITA would have recognized that this trigger-based architecture ensures the social media portion remains dormant until deliberately engaged by user action. The social media portion is further controllable because it allows the user to add a comment to the feed. EX1006, 30:55-60 (“[P]ost comments related to the video and/or the at least one interactive object in the video that are displayed when the overlay that tracks the interactive object is clicked or touched.”), 14:51-58, 28:44-50. This posting capability extends user control beyond mere viewing to active participation in the social dialogue. Users exercise discretion over both when to access the social media portion and what content to contribute to the object-specific discussion threads. Therefore, through both activation control and content contribution mechanisms, Ohliger establishes that the social media portion within the second interactive layer is fully *controllable by the user*, enabling selective engagement with social features based on individual viewing preferences.

144. Furthermore, as shown in Figure 26, the social media portion is controllable because it can be closed or exited by the user. This closing functionality provides users with complete discretion over the visibility of social content. A POSITA would have recognized that the “X” button visible in the interface

represents a standard dismissal control that removes the social overlay from the viewing area. Thus, the *social media portion* is *controllable by the user* through the following: displaying it when the viewer's selects an object in the video, interacting with it by the viewer posting a comment, and closing it by the viewer clicking the "X" button. These three control mechanisms—activation through object selection, participation through comment posting, and dismissal through the close button—demonstrate comprehensive user control over the social media functionality. Therefore, Ohliger's implementation ensures that viewers maintain autonomy over their engagement with the *social media portion*, allowing them to initiate, participate in, or terminate social interactions at will during video playback.



145. “so as to interact with the at least one guest that is also a viewer of the video media through an interface displayed in the second interactive video layer via the second media playback device”: Ohliger discloses that comments between viewers (*i.e.*, **user** and **guest**) are a form of interaction. EX1006, 11:11-14 (“[G]iving video viewers the ability to comment and communicate with each other and with the creator of the video”). This multi-party communication system enables real-time social engagement across distributed viewing sessions. Because the other commenters are also described as “viewers,” a POSITA would have understood that the other viewers interact with the first viewer via a similar interactive layer on their

respective device. The symmetrical nature of the viewing experience ensures all participants access identical interactive features.

146. Further, Ohliger discloses the viewer devices including personal devices such as mobile phones, which a POSITA would have recognized as indicating that each viewer views the video via their own playback device. EX1006, 22:28-33 (“[T]he viewer 124 uses a computer, mobile phone or another electronic device ... to access the video content”), Fig. 3. These individual devices enable personalized viewing experiences while maintaining shared social connectivity. A POSITA would have understood that the viewer devices would have also included personal devices such as smartphones. EX1006, Fig. 13.

147. Additionally, a POSITA would have recognized that the method of interaction by each viewer is indicative of viewing and interacting with the content on a personal device such as a smartphone, thus, a *second media playback device*. EX1006, 22:38-42 (“The viewer 124 interacts with the video overlay by selecting or rolling over a region representing a video object using a conventional pointing device 218, and/or using a touch sensitive screen 214 such as is known in the art.”). The touch-based interaction paradigm aligns with mobile device capabilities. It was well known at the time that smartphones operated primarily through touch sensitive screens. Therefore, Ohliger’s disclosure of touch-based interaction combined with mobile phone usage establishes that multiple viewers engage with the *social media*

*portion* through their individual *media playback devices*, enabling distributed yet connected social experiences.

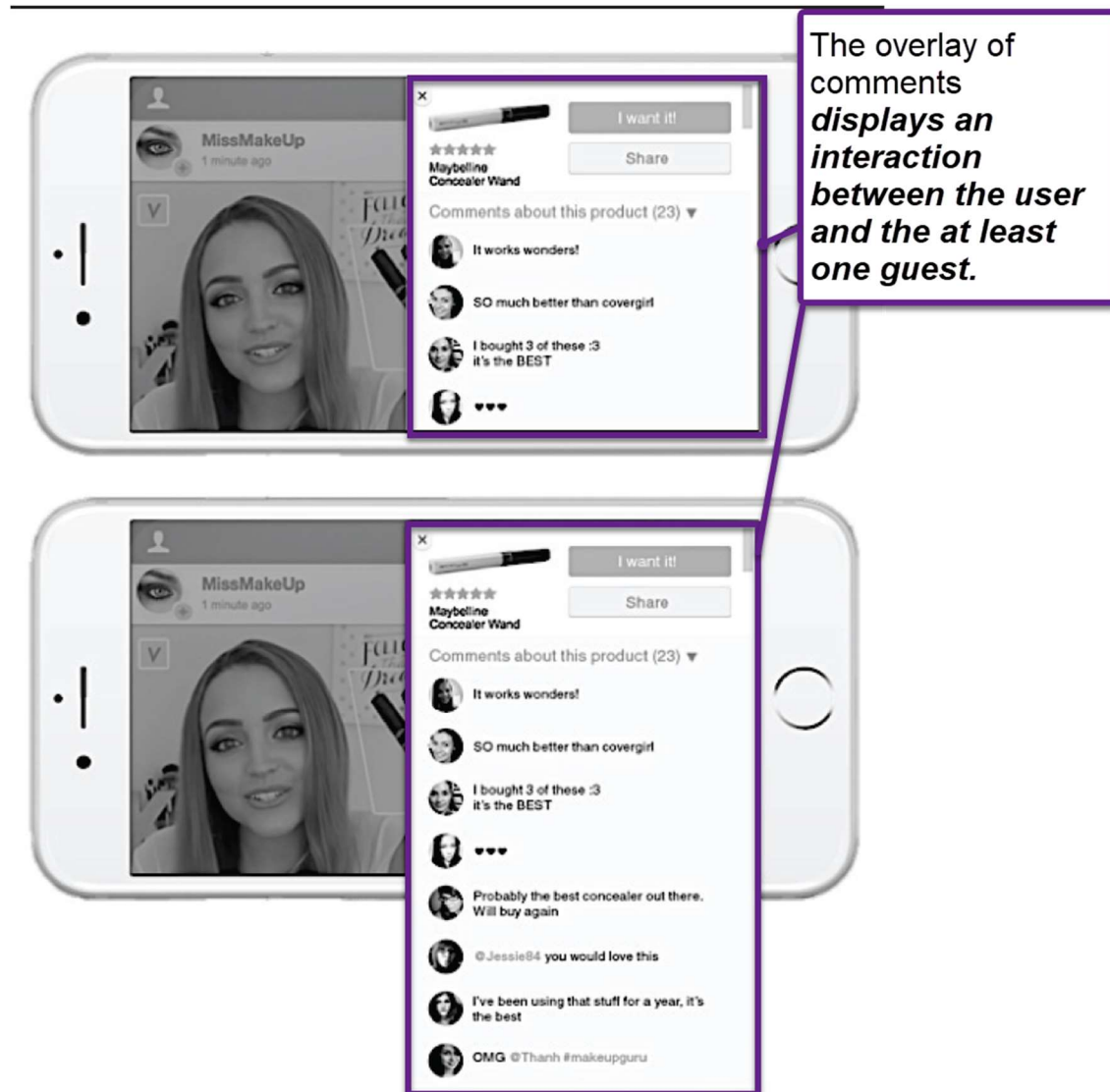
- (xi) **1[c][iii]: wherein the second interactive video layer displays an interaction between the user and the at least one guest, and past interaction activity between a plurality of users in the social media portion, and**

148. In my opinion, Ohliger discloses this limitation. Ohliger's overlay of comments on the video displays current and past interactions between multiple viewers (*i.e. the user and guest*). EX1006, 11:11-14 ("The invention also includes giving video viewers the ability to comment and *communicate with each other* and with the creator of the video in regards to a specific interactive object in the video"). The system preserves and presents these communications as persistent conversation threads associated with specific objects.

149. As shown in Figure 26, the comments overlay displays interactions between viewers. These visual representations of multi-party conversations demonstrate the social exchange occurring within the platform. The interface presents these interactions in a chronological or threaded format that allows viewers to follow the conversation flow. A POSITA would have understood that this display mechanism enables asynchronous communication, where viewers can read previous comments and add their own contributions regardless of when other participants engaged with the content. Therefore, through the persistent display of viewer

comments and communications, Ohliger's system enables *the user and guest* to engage in ongoing dialogues about specific objects, with the overlay serving as the visual record of these interactions across multiple viewing sessions.

Fig. 26  
Share & Comment



150. The displayed comments also include past interactions between users.

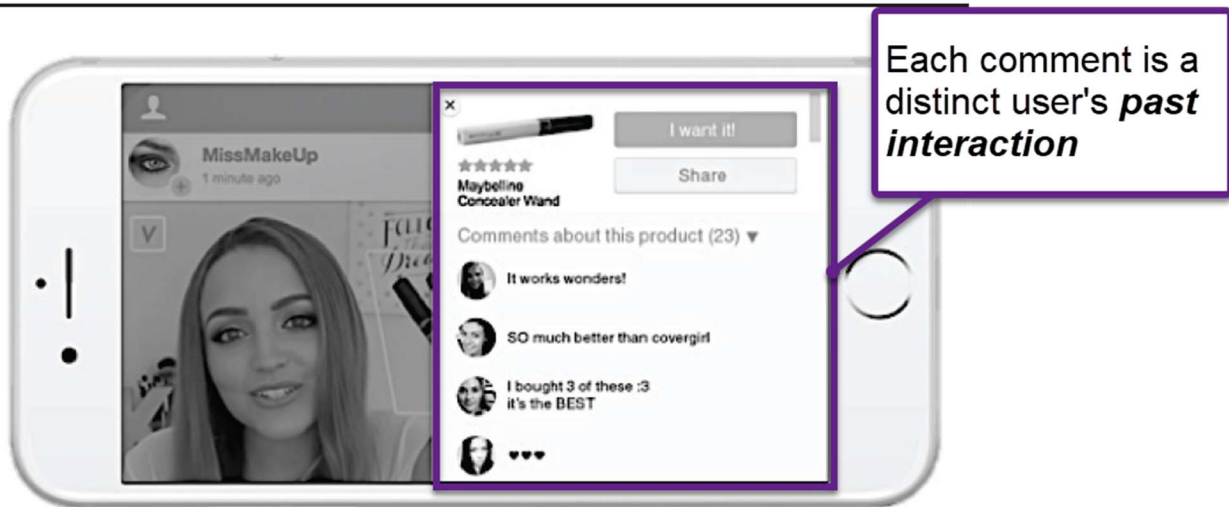
EX1006, 22:42-45, 14:51-59, 24:46-57. This temporal persistence enables viewers

to observe conversation histories associated with each object. Figure 26 displays comments from different viewers each with unique profile picture alongside each comment. The visual differentiation through profile pictures allows viewers to identify distinct participants in the conversation threads. These comments were added by several distinct viewers of the video (*a plurality of users*), including past comments that were already posted by the time the overlay with the comments is first shown (*past interaction activity*).

151. The presentation of both contemporaneous and historical comments creates a cumulative social record for each object. Through this mechanism, Ohliger demonstrates that the social media portion captures and displays *past interaction activity* from *a plurality of users*, enabling new viewers to observe and participate in established conversations around specific objects. Therefore, the system provides continuity of social engagement across multiple viewing sessions, with each viewer able to access the accumulated commentary from *past interaction activity* between *the user and guest*.

## Fig. 26

### Share & Comment



(xii) 1[c][iv]: wherein the user is capable of maintaining and toggling between a plurality of different conversations displayable in the social media portion in the second interactive video layer as an overlay to the video media displayed by the first media playback device,

152. In my opinion, Ohliger discloses this limitation. In Ohliger, each interactive video object enables an overlay of a distinct set of viewer comments related to the particular video object. EX1006, Fig. 26. This object-specific commenting architecture creates separate conversation threads for individual items within the video. Since each interactive video may contain more than one interactive video object, the viewer may *maintain and toggle between different conversations* for each video object by selecting one video object and then subsequently selecting another video object. EX1006, Fig. 18, Fig. 27. A POSITA would have recognized

that this toggling mechanism allows viewers to navigate between distinct social discussions without losing the context of each object-specific conversation.

153. As discussed regarding limitations 1[c][i]-[iii], a viewer or creator (*i.e. the user*) may “share and view sub-comments related to individual products.” EX1006, 29:15-16. These product-specific discussions remain visually associated with their corresponding objects. The comments are superimposed over and cover a portion of the video (*i.e. displayable ... as an overlay to the video media*). See EX1006, Fig. 26; *see also* Ohliger at 24:46-57 (“Once the object has been selected ... the viewer displays an abstract of the video object in **the form of an overlay** ... includ[ing] descriptive information ... about the object ... like a comment about the object ... [or] comments by other viewers.”).

154. Ohliger further describes distinguishing between each of multiple video objects appearing in a single video. EX1006, 19:18-22, *see also* EX1006, 28:46-50, 29:5-10, 29:17-18. This differentiation enables the system to maintain separate conversation contexts for each identified object. The viewer or creator (*i.e. the user*) can then view and add comments to each distinct video object in a particular video, also referred to as “sub-comments.” EX1006, 29:15-16.

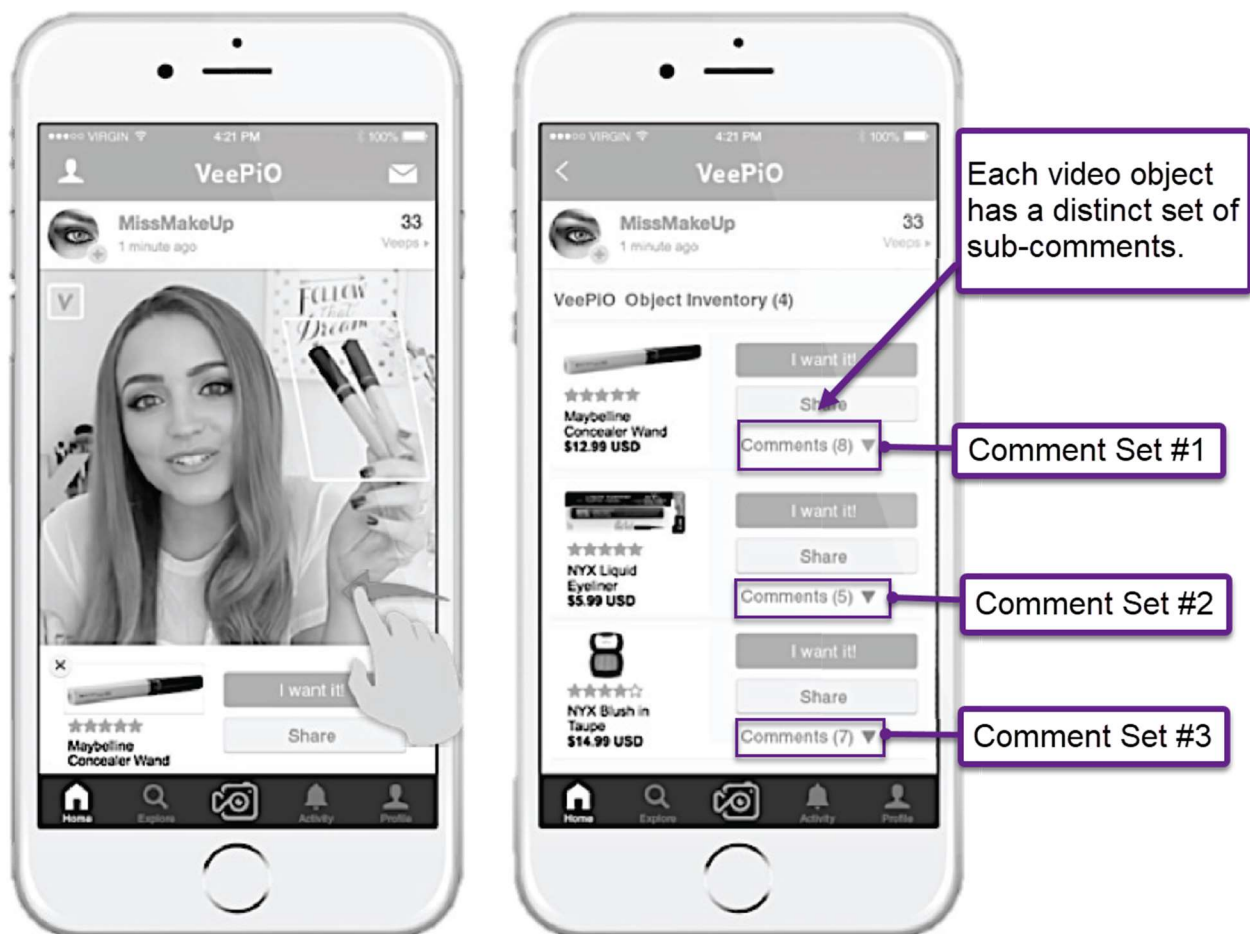
155. Figure 26 shows a single set of sub-comments for a particular video object, and Figure 27 shows that each video object has a distinct set of sub-comments (*i.e., a plurality of different conversations*) as indicated by the number of comments

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

for each. A POSITA would have understood that the numerical indicators visible in Figure 27 represent comment counts for individual objects, demonstrating that the system maintains separate conversation threads. Therefore, through this multi-conversation architecture, Ohliger enables *the user* to *maintain and toggle between a plurality of different conversations*, each associated with a specific object and *displayable ... as an overlay to the video media*.

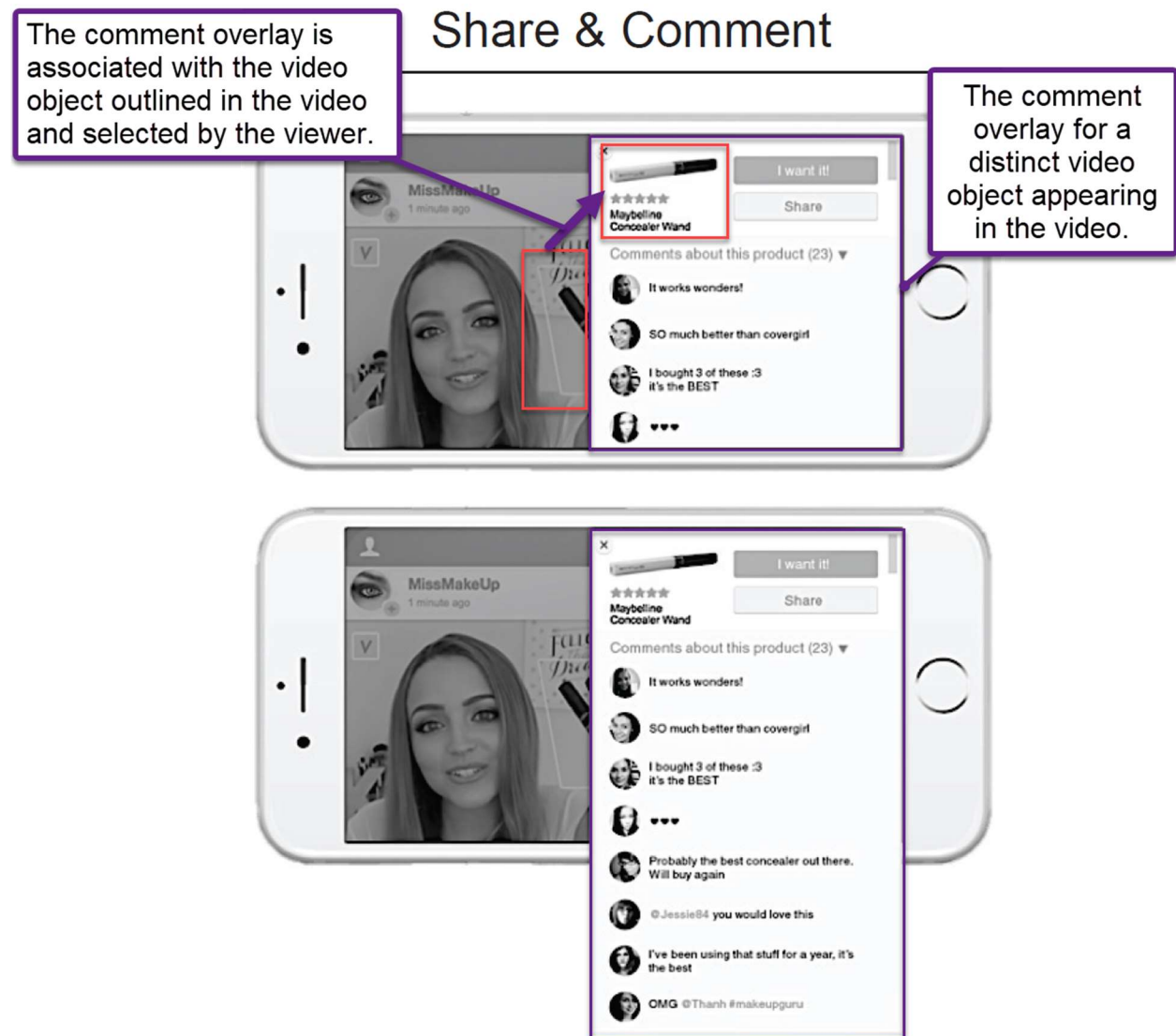
Fig. 27  
Veep Inventory



156. Therefore, viewers may click on different video objects to *maintain and toggle between different conversations* associated with their respective video objects. This selection-based navigation enables seamless transitions between distinct discussion threads while maintaining the continuity of each conversation. Each conversation is displayed as overlays within the same portion of interactive layers, *the social media portion*. A POSITA would have recognized that this unified display location ensures consistent user experience while preserving the discrete nature of each object's conversation thread.

157. For example, upon selecting the "Maybelline Concealer Wand" in Figure 27, conversations for that object are displayed as shown in Figure 26 (also annotated as Comment Set #1 in Figure 27 above). The visual transition from object selection to comment display demonstrates the direct relationship between object interaction and conversation access. Through this implementation, Ohliger establishes that *the user can maintain and toggle between different conversations* by selecting different objects, with each conversation thread remaining accessible through its associated visual overlay within *the social media portion*.

Fig. 26

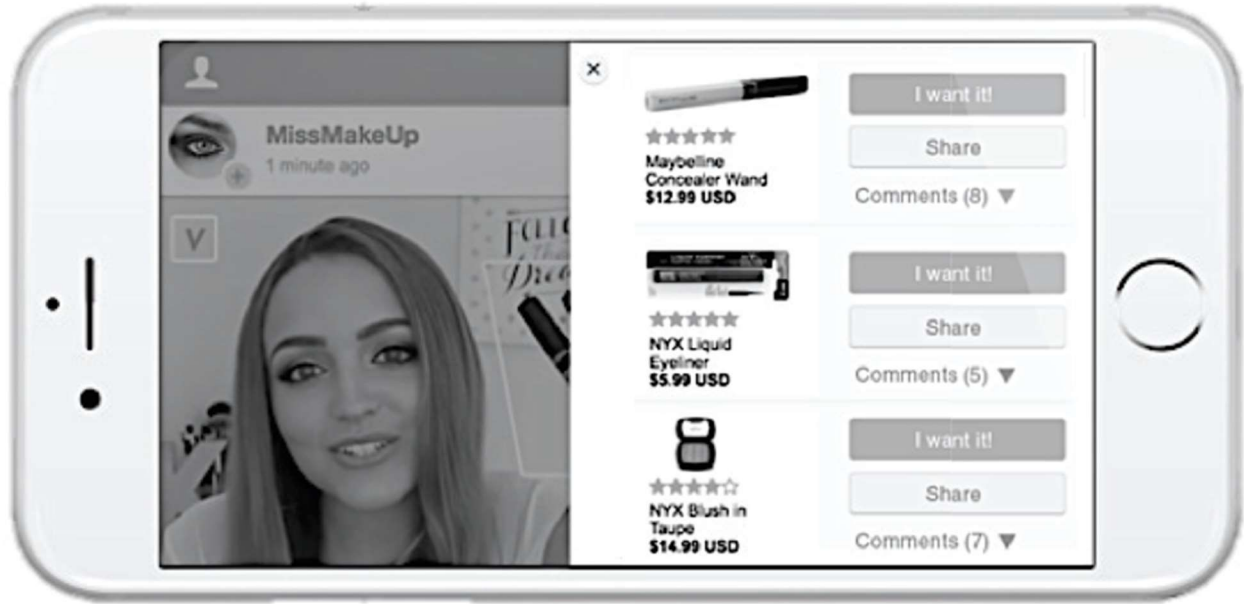


158. Alternatively, Ohliger at least renders obvious *maintaining and toggling between a plurality of different conversations* through overlaying the “Object Inventory” screen of Figure 27 on the video. This alternative implementation would extend the existing overlay architecture to display multiple conversation through the same overlay. A POSITA would have understood that each numbered indicator, *e.g.* “Comments (8),” in Figure 27 represents a user control to

display the associated conversation thread as similarly expanded to display the full conversation as in Figure 26. A POSITA would have recognized that this approach provides viewers with a comprehensive view of all available object-specific discussions within a single interface.

159. Ohliger already discloses that a user may hit the “V” button to highlight all video objects in a particular frame. This existing functionality demonstrates the system’s capability to visually indicate multiple interactive elements concurrently. In my opinion, a POSITA would have been similarly motivated to display the additional information shown in Figure 27 (including the different sub-comments) in the same manner. The technical implementation would involve extending the existing overlay generation mechanisms to incorporate the inventory view directly onto the video playback.

160. A POSITA would have understood that overlaying the object inventory would provide several technical advantages. The streamlined display of multiple conversations would enable rapid navigation between different object discussions without requiring sequential object selection. This modification would leverage Ohliger’s existing overlay infrastructure while enhancing the user’s ability to survey and access multiple conversation threads. A mock-up of the resultant combination is shown below:



161. In my opinion, a POSITA would have recognized this inventory overlay would benefit viewers' comparison shopping experiences by allowing viewers to view information on different products more efficiently. Ohliger discloses that splitting the screen real estate up to allow for the simultaneous display of distinct user interfaces for functionalities enables viewers "to browse that product while they watch it in action." EX1006, 29:19-21, Fig. 28. This establishes Ohliger's recognition of the value of concurrent information presentation.

162. A POSITA would have recognized that the overlay of the inventory to link the viewer to multiple conversations provides this well-known benefit. The display of multiple conversation indicators would streamline user navigation. A POSITA would have understood that by applying similar overlays to sets of sub-comments, users could then readily select relevant sub-comments for display,

thereby *maintain[ing] and toggling between ... different conversations* as already described in connection with Fig. 27.

163. Indeed, this would have required displaying UIs already disclosed in Ohliger (Fig. 27) in a form already disclosed in Ohliger (Fig. 26) in response to actions already disclosed in Ohliger (Fig. 23; 29:5-10). By providing this overlay, the POSITA would have streamlined the user's experience from requiring a succession of operations (*e.g.* selecting an object, viewing the comments, closing the object's overlay, and then selecting another object) to a single operation (selecting the object from the object inventory screen triggering the display of that object's comment thread and then selecting a second object from the object inventory overlay to trigger the display of that second object's comment thread). This reduction in operational steps represents a clear improvement in user interface efficiency.

164. In my opinion, a POSITA would have had reasonable expectation of success in doing so because it would have entailed combinations of two existing user interfaces, both screens of Figure 27, in overlays patterned after other overlays disclosed in Ohliger, *e.g.* Figure 18, 26, 28. The technical feasibility stems from applying Ohliger's established overlay techniques to combine already-disclosed interface elements. A POSITA would have recognized that this modification requires only routine implementation skills, combining known UI components using Ohliger's existing overlay infrastructure to achieve predictable results.

(xiii) 1[c][v]: wherein the plurality of users either includes the user or does not include the user; and

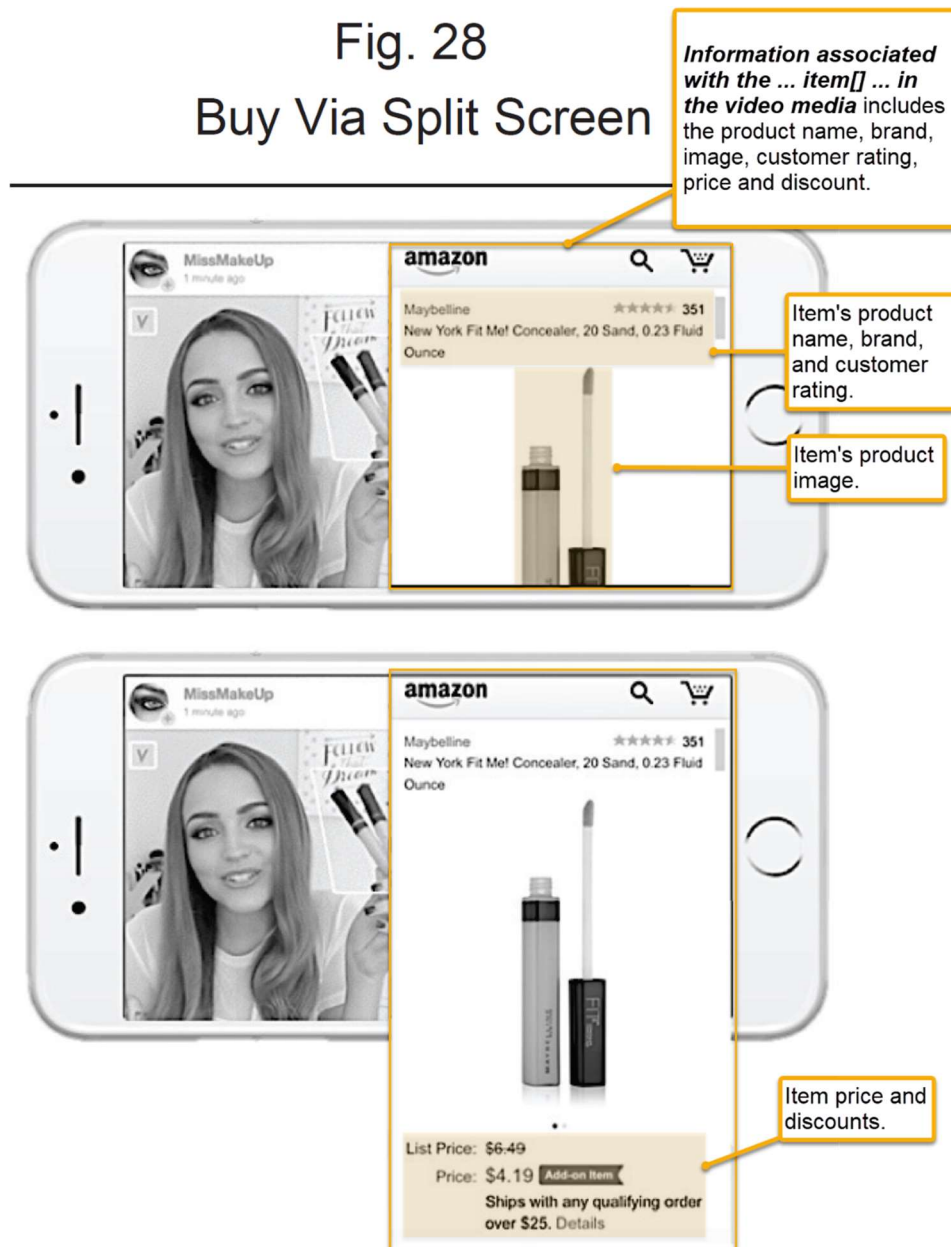
165. In my opinion, Ohliger discloses this limitation. Ohliger discloses that viewers or creators (*the user*) may add comments to video objects. Therefore, *the plurality of users ... includes the user*. And if users have not left comments, then *the plurality of users ... does not include the user*. Either way, this limitation is met.

(xiv) 1[d]: d) distributing information associated with the one or more items, services, or songs in the video media to the user.

166. In my opinion, Ohliger discloses this limitation. As discussed for limitation 1[c] – 1[c][iv], Ohliger discloses overlaying hyperlinks, the product name, brand, image, customer rating, price, discount, and comments (*i.e., information associated with the one or more items ... in the video media*) on the video media during playback on the user’s device (*i.e., to the user*), as seen in Figure 28 below. *See also* EX1006, 22:7-11; limitations 1[pre], 1[b][ii] (additional analysis of claimed *information*). The overlaid information encompasses multiple data categories for each identified object. A POSITA would have recognized that this overlaid information transforms video objects into actionable shopping opportunities. The overlay mechanism delivers product specifications and commercial details directly within the video viewing interface. Through displaying price points, discount offers, and customer ratings alongside the visual representation of the product in use,

Ohliger's system provides contextual purchasing information at the moment of viewer engagement. Therefore, by overlaying this array of *information associated with the one or more items ... in the video media* during playback, Ohliger enables delivery of comprehensive product data *to the user* through their media playback device.

Fig. 28  
Buy Via Split Screen



167. Moreover, Ohliger discloses that the **distributing** of limitation 1[d] is performed after the **generating** of limitation 1[c]. The temporal sequence of operations ensures proper data preparation before transmission. For example, Figure 1's process 750 describes "[s]erv[ing] object information to interactive video player." EX1006 Fig. 1. This serving step occurs subsequent to the overlay generation and linking processes described in earlier steps.

168. A POSITA would have recognized that this sequential ordering reflects the technical necessity of completing data generation before distribution can occur. The system must first process the video content, create the interactive overlays, and compile the associated information before transmitting the integrated package to viewers. Therefore after the **generating** steps of limitation 1[c], the **distributing** is performed with the "serv[ing]" of the **video data**. EX1006 Fig. 1. This ordering ensures that viewers receive fully processed interactive content with all overlay elements properly positioned and linked to their corresponding objects. Through this sequence from generation to distribution, Ohliger demonstrates the complete data flow from server-side processing to client-side delivery of the interactive video experience.

169. Finally, PO distinguishes "**items**" and "**items ... of interest**," contending that only the "**items**" have to be "in" the video media, while the "**items ... of interest**" are part of the "**first interactive video layer**" in the form of "link[s]"

to the marketing and/or purchasing information.” EX1017, 6-9. Even if such separate “items” and “items ... of interest” are required, they are disclosed by Ohliger, where the video objects identified at step 700 (*items*) are within the video itself and visual outlines created and linked to winning bidders at step 745 (*items ... of interest*) are part of the video overlays. *See, e.g.*, Ohliger, 11:54-57, 12:38-41; *see also* limitations 1[a], 1[b][i], 1[b][iv].

**(b) Claim 2**

**(i) The method of claim 1, further comprising:**

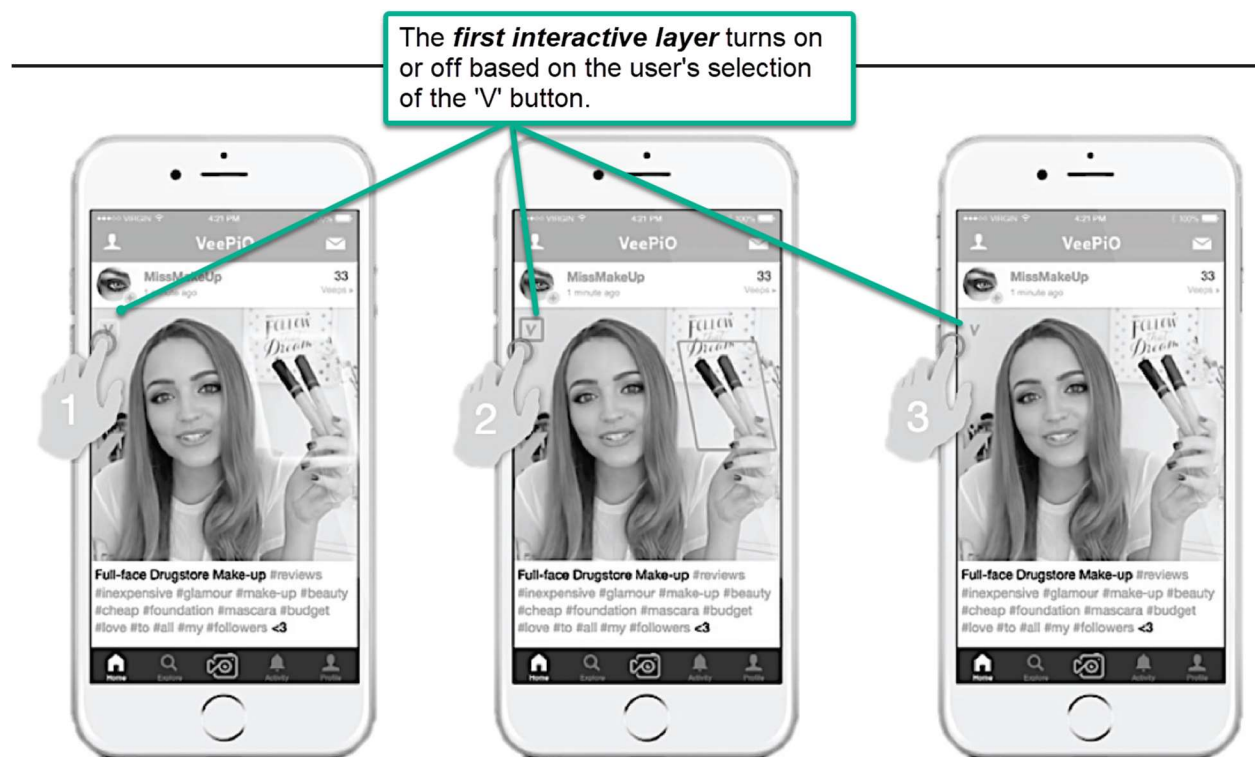
**2[a][i]: enabling the user to present the video data with or without the first interactive video layer being turned on or off based on selection by the user to view or not view the first interactive video layer, respectively; and**

170. In my opinion, Ohliger discloses this limitation. Ohliger discloses enabling a viewer or creator (*the user*) *to turn on or off* the video object box overlays (*first interactive video layer*) by toggling a “V” button in the user interface. EX1006, 29:5-10 (“Once published, the video shows up first in the feed. Tapping the ‘V’ once reveals a transparent halo around each Veep’d object. (1st Screen) Tapping twice places colored boxes on each object so users can view multiple objects with ease. (2nd Screen) Veep’ers can turn objects off by tapping the ‘V’ watermark a third time. (3rd Screen).”), 29:11-12.

171. The multi-state toggle mechanism provides control over the overlay’s visibility. A POSITA would have recognized that this three-tap sequence cycles

through distinct display states: no overlay, transparent halos, colored boxes, and back to no overlay. Each state alters the interactive elements. The implementation allows viewers to customize their viewing experience based on their immediate needs—whether focusing on the video content without distraction or actively engaging with the shopping features.

172. Thus, the “V” button allows a user *to select whether to view or not view the first interactive video layer*. Through this toggle control, Ohliger establishes user control over the display of interactive overlays, enabling viewers to activate or deactivate the *first interactive video layer* at will during video playback. The persistence of this control throughout the viewing session ensures that users maintain discretion over their interaction with the overlay elements throughout the video.



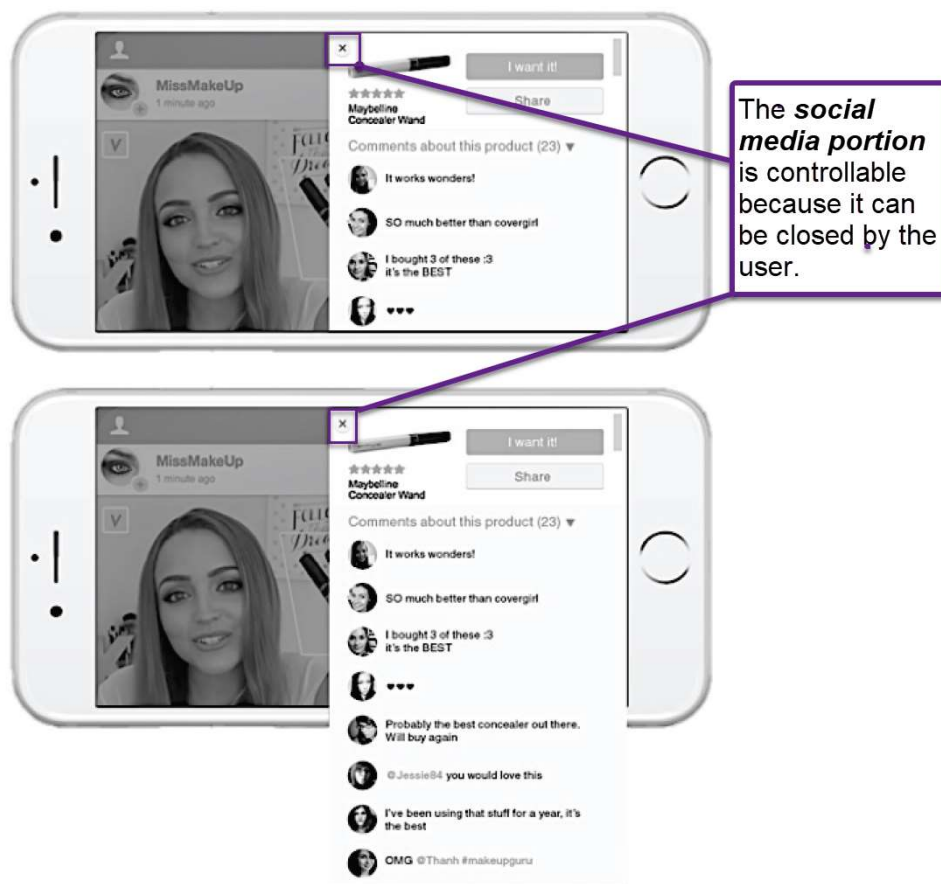
173. Additionally, Ohliger also describes *turn[ing] on or off based on selection by the user to view or not view the first interactive video layer* as part of step 760 to the process diagram of Figure 1: “[i]n step 760, a video viewer interacts with the video overlay by rolling-over or selecting a video region thereby triggering the display of relevant content associated with the video object.” EX1006, 12:42-45; *see also* EX1006, 12:45-50.

174. This user-initiated triggering mechanism demonstrates selective overlay activation based on viewer intent. A POSITA would have recognized that roll-over and selection actions represent deliberate choices to engage with the interactive layer. The system responds by either displaying or hiding overlay content, effectively allowing viewer control over the *first interactive video layer* visibility.

- (ii) **2[a][ii]: enabling the user to present the video data with or without the second interactive video layer being turned on or off based on selection by the user to view or not view the second interactive video layer, respectively.**

175. In my opinion, Ohliger discloses this limitation. Ohliger discloses that the user comments overlay (*second interactive video overlay*) associated with a particular item in a video is displayed by the viewer interacting with the associated outline overlay around the item. EX1006, Fig. 18, Fig. 26. The interdependence between the first and second layers creates a cascading control mechanism.

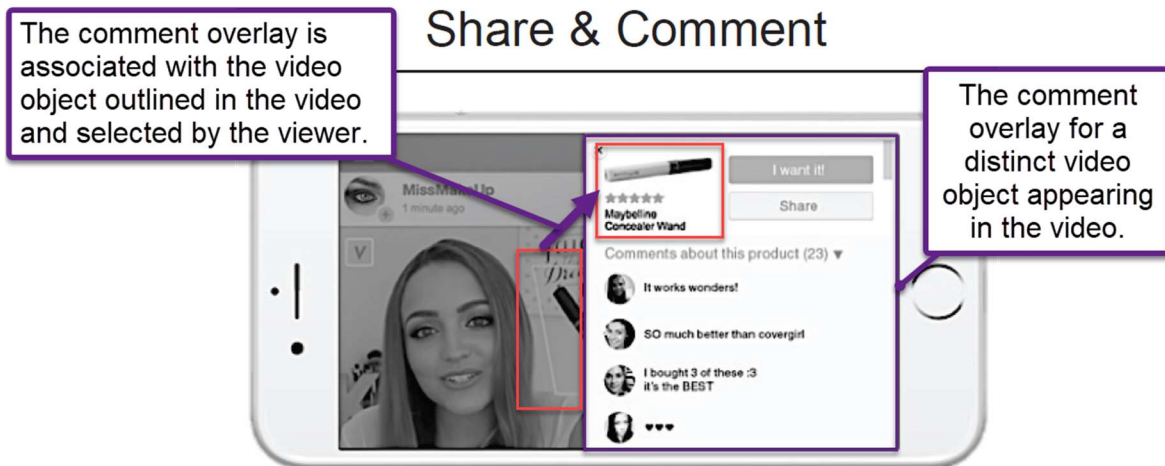
Accordingly, toggling on or off the box overlay via the “V” button also enables the user to *turn on or off* the user comments overlay based on user selection of the “V” button. See limitation 2[a][i], above, EX1006, 29:5-12. A POSITA would have recognized that disabling the first layer necessarily prevents access to the second layer, as the interactive triggers become unavailable. Additionally, Ohliger discloses an “X” close button on the comments overlay in both Figure 18 and Figure 26. A POSITA would have recognized that this allowed for *turn[ing] ... off* the comments overlay (*the second interactive video layer*).



176. Moreover, Patent Owner's Infringement Contentions accuse the functionality of "[c]hoos[ing] who can comment on ... videos," and highlights the distinction between videos with comments overlaid and not overlaid. EX1005, 58-65. This distinction reinforces the controllable nature of the comment overlay display. Similarly, Ohliger discloses comment overlays that appear only when the viewer interacts with the associated item in the video as seen in Figure 26. Therefore, through both the "V" button toggle and interaction-based display, Ohliger enables *the user* to control whether the *second interactive video overlay* appears during video playback.

Fig. 26

## Share & Comment



(c) **Claim 3**

(i) **The method of claim 2, further comprising:**

**3[a][i]: displaying the video media and the first interactive video layer to the user; and**

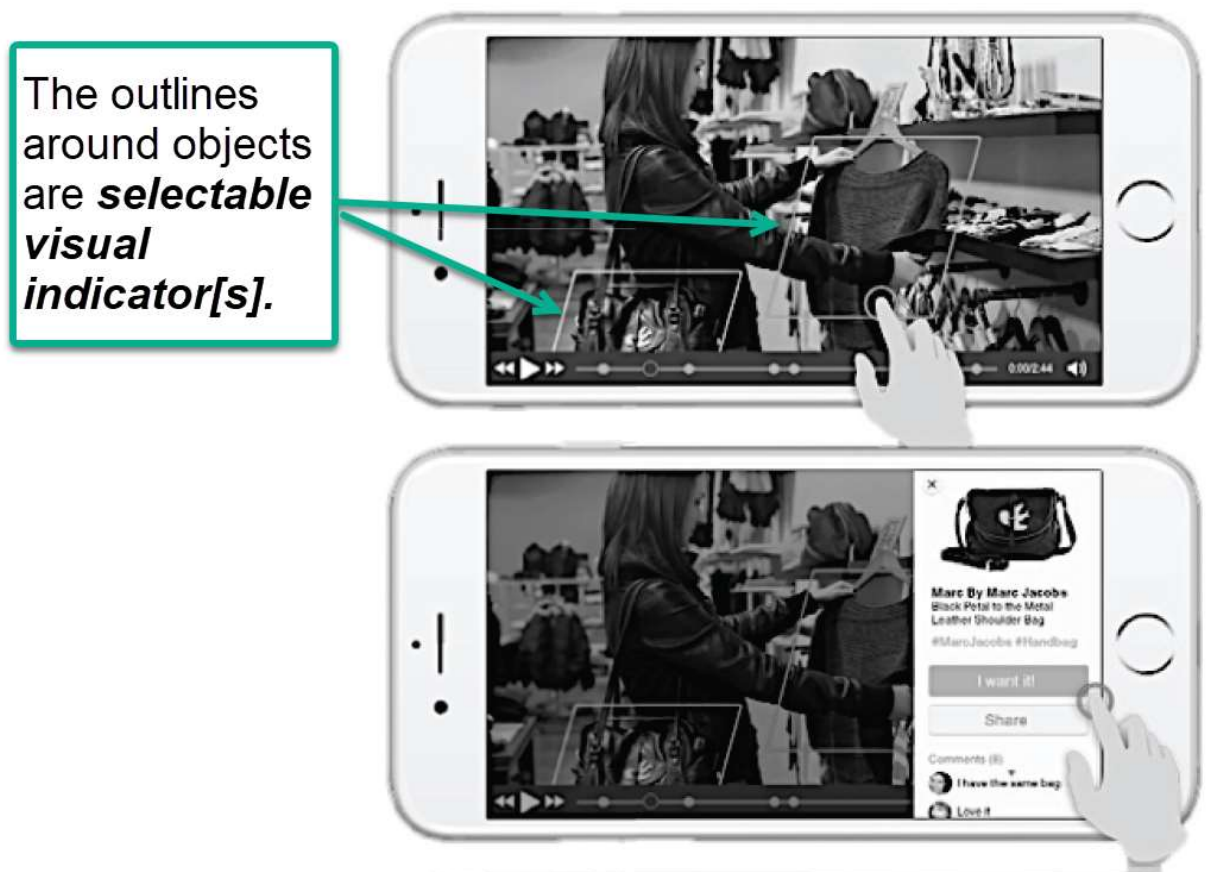
177. In my opinion, Ohliger discloses this limitation. This limitation is identical in scope to limitation 1[b][iv]’s *displaying the video media and the first interactive video layer* and therefore is disclosed by Ohliger for the same reasons.

(ii) **3[a][ii]: receiving a selection of a given selectable visual indicator of a given item, service, or song of interest from the user to display marketing information of the given item, song, service, or song of interest to the user.**

178. In my opinion, Ohliger discloses this limitation. Ohliger describes overlaying a “selectable hyperlink whose position tracks a dynamic location of the video object in the video.” EX1006, 3:48-53, 8:61-9:2. This dynamic tracking ensures the selectable region remains aligned with the object throughout its appearance in the video. The viewer (*user*) makes a *selection* through “clicking or touching area around objects in the video.” EX1006, 9:5-11. A POSITA would have recognized that these interaction methods represent standard input mechanisms for selecting visual interface elements.

179. Figure 18 illustrates the user’s *selection of a given visual indicator of a given item ... of interest* in the first screen where the hand gesture graphic indicates the user’s selection via the smartphone’s touch screen. The visual feedback provided

by the hand cursor confirms the interactive nature of the overlay elements. Through this touch-based selection mechanism, Ohliger demonstrates how *the user* engages with specific visual indicators to access information about items of interest. Therefore, Ohliger's implementation of selectable, dynamically-tracked overlays that respond to user touch or click input fully discloses the *selection of a given visual indicator of a given item ... of interest by the user.*



180. Ohliger discloses *receiving a selection ... to display marketing information* because “[o]nce the object has been selected ... the viewer displays an abstract of the video object in the form of an overlay on top of the current viewing

window ... includ[ing] descriptive information ... about the object (name, brand name, make, model, vendor, price, etc)” (*marketing information*). EX1006, 24:46-57. The system’s response to user selection triggers the immediate presentation of commercial product data.

181. Figure 18 illustrates the *receiving a selection* with the screen displaying the overlay popup containing information associated with a previously selected object, *i.e.* the purse. The transition from selection to information display demonstrates the system’s processing of the user’s input request. As discussed above, the *marketing information* includes the name, brand, and product image as seen in the exemplary overlay in Figure 18. A POSITA would have recognized that this comprehensive product information—encompassing identification, branding, and pricing details—constitutes *marketing information* designed to facilitate purchasing decisions. Therefore, Ohliger’s disclosure of selection-triggered display of product descriptors, vendor data, and pricing establishes the *receiving a selection ... to display marketing information* for the selected item of interest.

**(d) Claim 4**

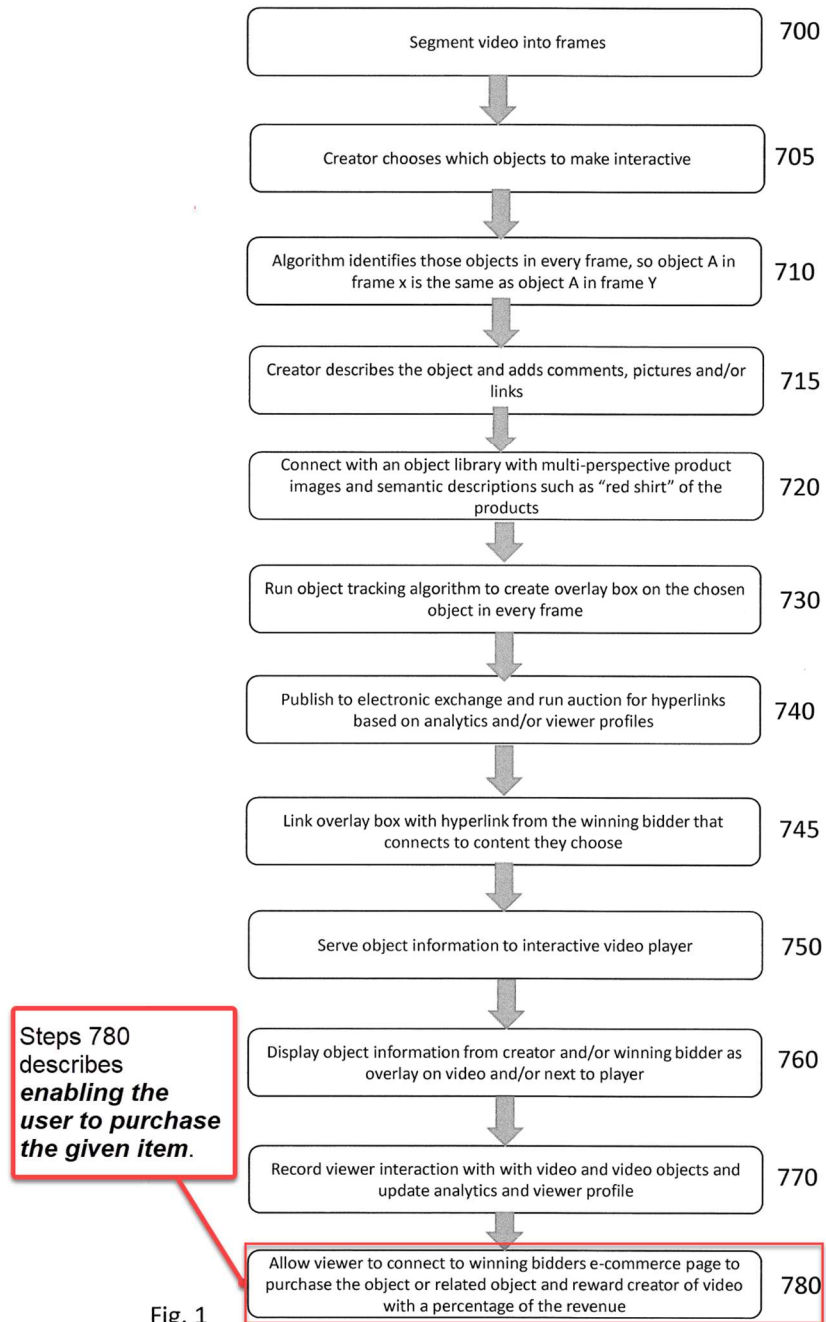
**(i) The method of claim 3, further comprising**

**4[a]: enabling the user to purchase the given item, song, or service of interest from the marketing information.**

182. In my opinion, Ohliger discloses this limitation. Ohliger integrates the video overlays with e-commerce functionality (*enabling the user to purchase the*

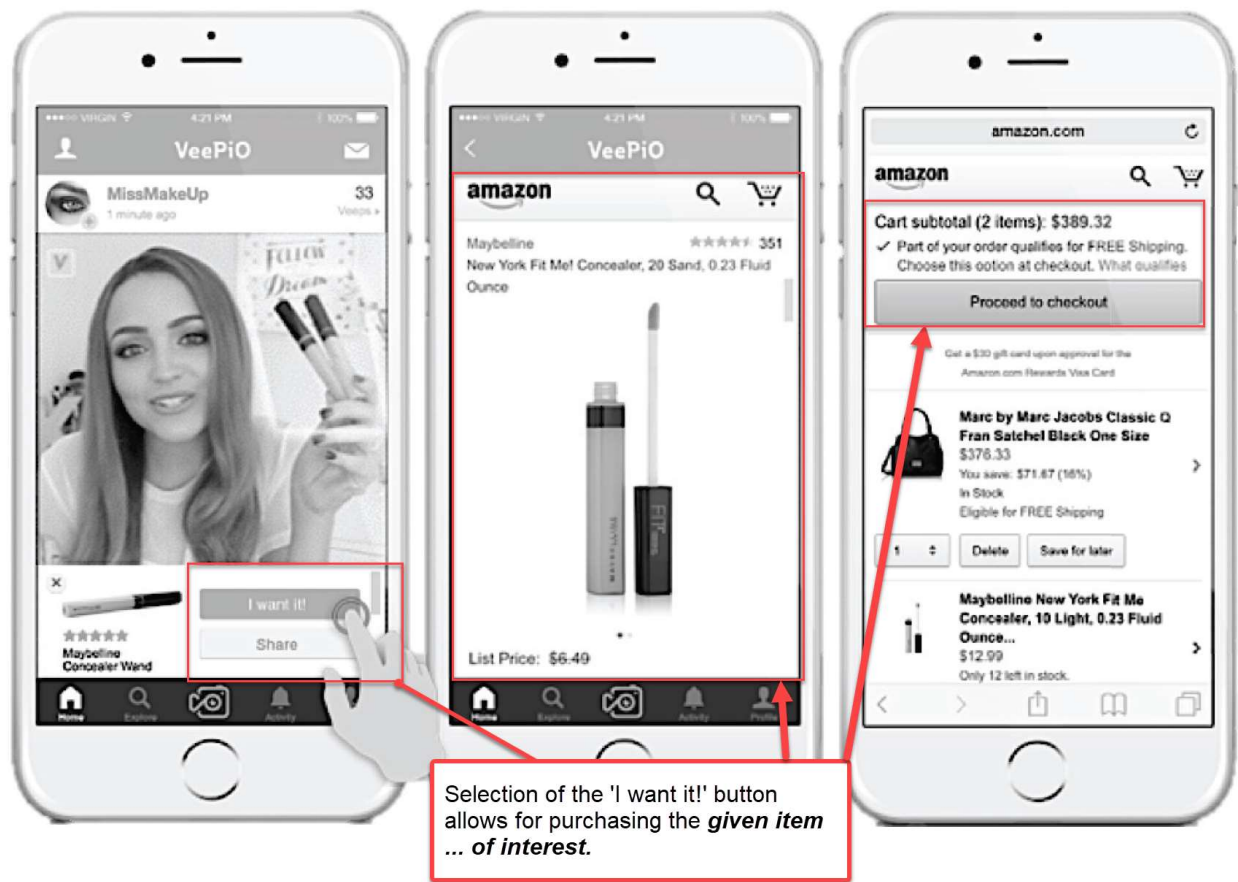
*given item ... of interest*) such as linking to “an online store where a product or products associated with the object in the video can be purchased.” EX1006, 3:58-62, 5:13-16. This direct connection between object selection and purchasing capability transforms video viewing into an interactive shopping experience.

183. Figure 1’s step 780 describes “[a]llow[ing] viewer to connect to winning bidders e-commerce page to purchase the object.” EX1006, Figure 1. The system facilitates seamless transitions from product discovery to transaction completion. Therefore, through the combination of selectable overlays linked to e-commerce pages, Ohliger’s system *enabl[es] the user to purchase the given item ... of interest* directly from the interactive video interface, completing the journey from visual identification to commercial transaction.



184. Figure 25 shows the viewer (*user*) purchasing the item by interacting with the product information overlaid on the video. EX1006, 29:13-15. This demonstration illustrates the complete e-commerce workflow from selection to transaction. The interface presents purchasing options directly within the video

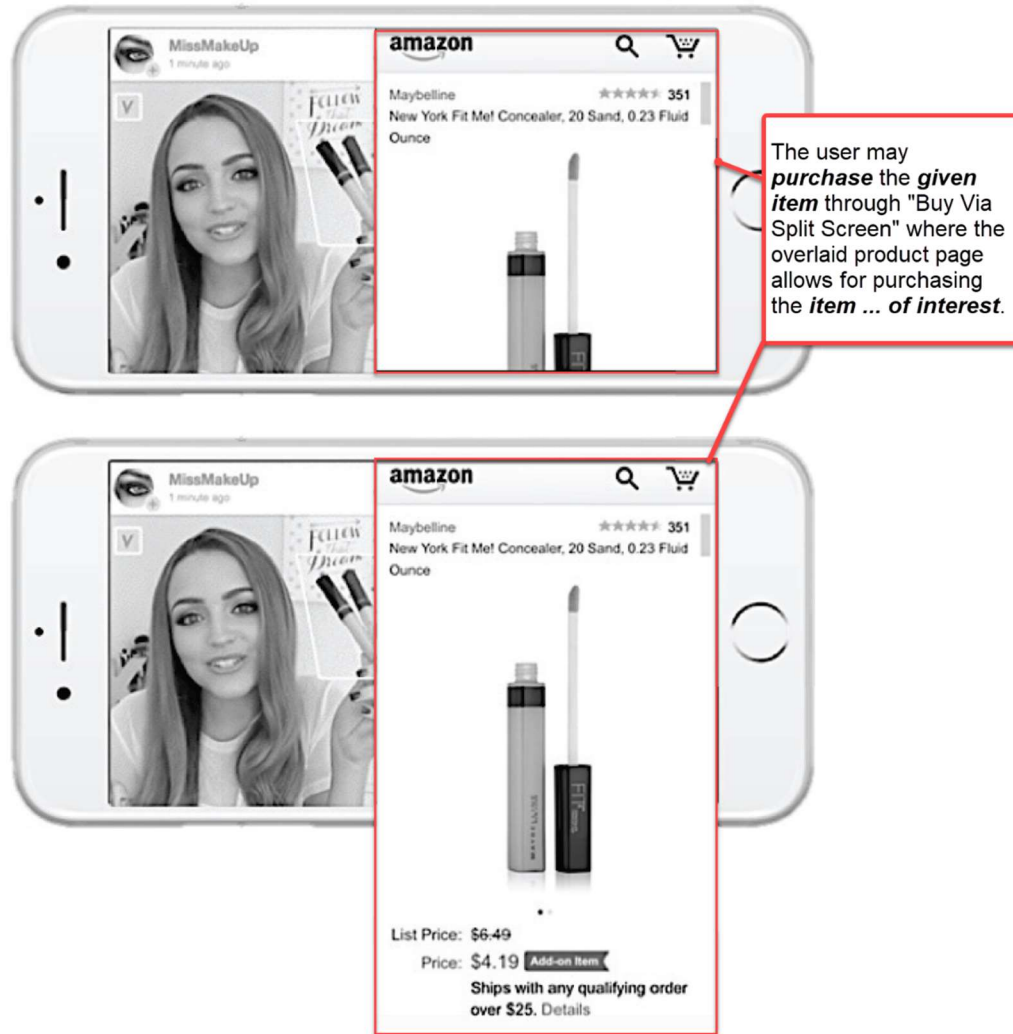
viewing environment, eliminating the need for external navigation. A POSITA would have recognized that this in-video purchasing capability streamlines the conversion from viewer interest to completed sale. Through the interactive elements shown in Figure 25, Ohliger establishes that the system *enables the user to purchase the given item ... of interest* without leaving the video playback interface.



185. Figure 28, titled “Buy Via Split Screen,” shows a viewer’s ability to purchase through the overlay of product information, “giving them a way to browse that product while they watch it in action.” EX1006, 29:19-21. This split-screen design, which is an e-commerce overlay (Amazon.com in Figure 28) on top of the

video itself and not an actual split or division of the screen, enables simultaneous video viewing and e-commerce browsing within a unified interface. A POSITA would have recognized that displaying the product in use while presenting purchasing options leverages the persuasive power of contextual demonstration. The implementation allows viewers to observe product performance while accessing commercial details, creating an integrated shopping experience. Therefore, Figure 28's split-screen purchasing interface confirms that Ohliger *enables the user to purchase the given item ... of interest* by providing direct e-commerce functionality overlaid on the active video content.

Fig. 28  
Buy Via Split Screen



(e) Claim 5

(i) The method of claim 4,

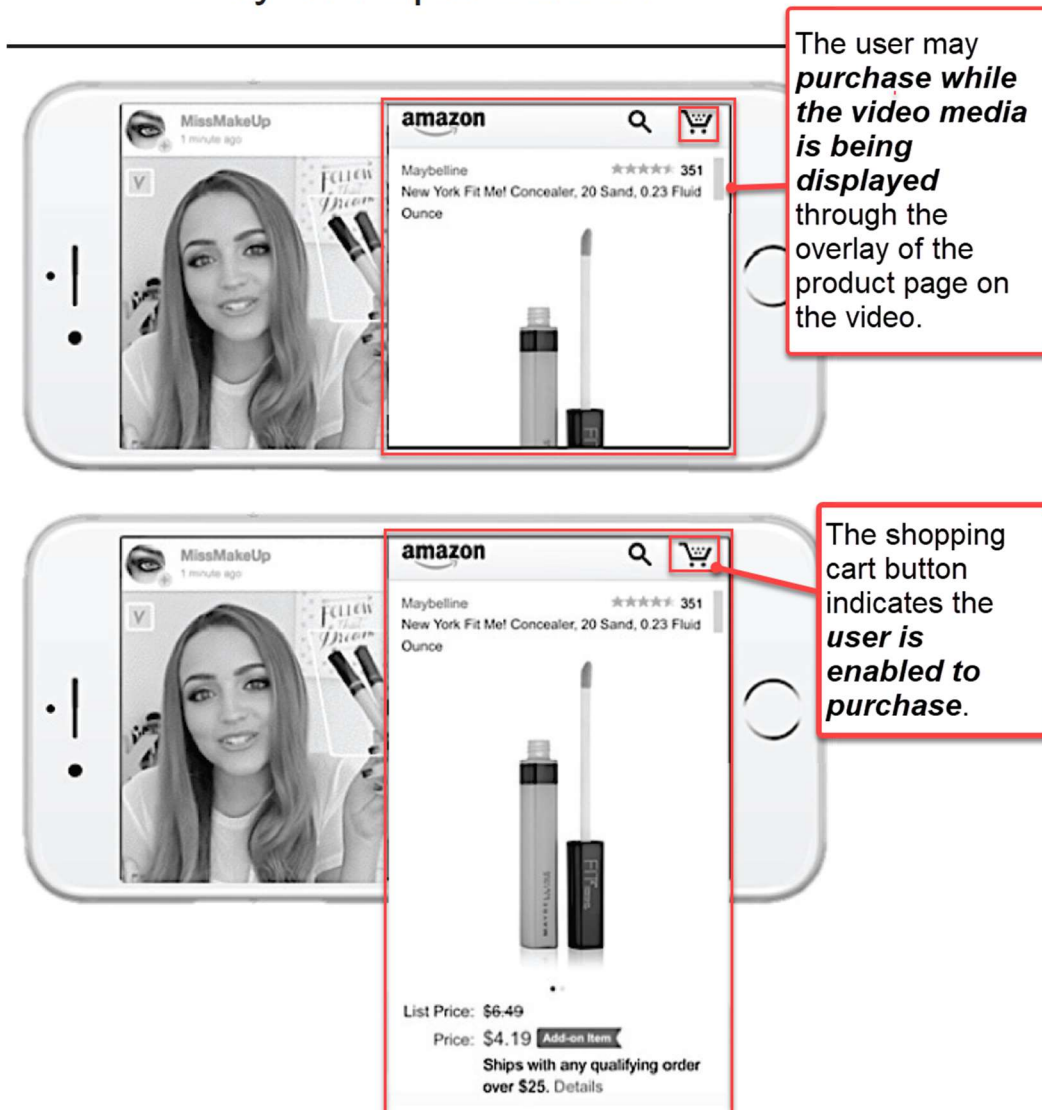
**5[a]: wherein the user is enabled to purchase while the video media is being displayed to the user.**

186. In my opinion, Ohliger discloses this limitation. Ohliger's overlays include purchasing functionality where a user can purchase the identified object

found in the displayed video. EX1006, 29:19-21 (“Users can view product pages via a split screen, giving them a way to browse that product while they watch it in action.”). The overlay of e-commerce functionality maintains visual context while enabling commercial transactions.

187. Figure 28 depicts that a user can “Buy Via Split Screen” the item. The interface demonstrates direct e-commerce integration within the video viewing environment. Additionally, the shopping cart icon on the product page illustrates the purchasing of the item in the overlaid page as annotated in Figure 28 below:

Fig. 28  
Buy Via Split Screen



188. A POSITA would have recognized that the shopping cart icon represents standard e-commerce functionality for adding items to a purchase queue. EX1006, 25:17-26 (“As described above, the user may be presented with the option to add the video object to a shopping cart”), 25:27-42 (“The shopping cart may be a conventional electronic shopping cart which enables the user to purchase the

object”). The presence of this purchasing mechanism within the overlay confirms that the system enables transaction completion without requiring navigation away from the video content. Therefore, through the overlay of product pages with embedded shopping cart functionality, Ohliger’s system *enables the user to purchase the given item ... of interest* directly through the interactive overlay interface while maintaining continuous access to the video demonstration of the product in use.

189. In addition, in my opinion, it would at least have been obvious to a POSITA to combine purchasing (as shown, for example, in Figure 25) into Figure 28’s “product page[]” as a means of streamlining the purchasing process for viewers of the video. The technical implementation would involve integrating established e-commerce functionality into the existing overlay architecture. A POSITA would have been motivated to streamline and reduce any friction against viewers purchasing objects in videos because it was one of the primary results from the system. EX1006, 24:17-20 (“the viewer selects one or more objects to purchase ... and ... the viewer completes the transaction.”), 25:19-22, 28:46-56.

190. Further, Ohliger discloses the system collecting revenue from sales of products in videos, thus, providing an economic incentive. EX1006, 28:51-56 (“VeePiO passes along 80% of the revenue to the video creator.”). This revenue-

sharing model creates alignment between platform success and creator compensation.

191. Ohliger recognizes the benefit of streamlining the purchasing process in a single app and view: “[u]sers can then purchase Veep’d objects straight from the video with no hurdles to checkout,” and “[n]o redirect is required in VeePIO ... giving them a way to browse that product while they watch it in action.” EX1006, 29:13-14, 29:19-21.

192. In my opinion, a POSITA would have had a reasonable expectation of success as implementing a purchase process into the product page was well within the skill of an ordinary artisan. The technical feasibility stems from combining disclosed elements. Further, Ohliger already disclosed the purchasing and e-commerce functionality with the overlay for browsing products. EX1006, Fig. 28. Therefore, integrating complete purchase capability into the overlaid product page represents a predictable variation that *enables the user to purchase the given item ... of interest* through streamlined, in-video commerce functionality.

**(f) Claim 6**

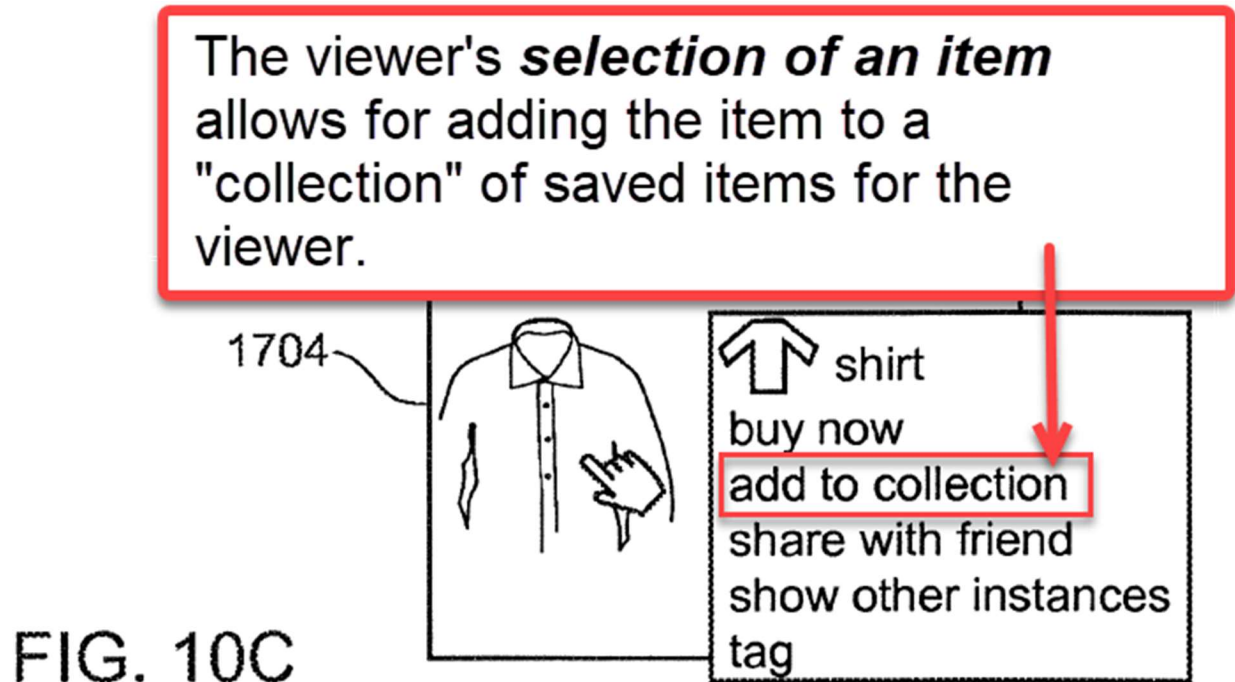
**(i) The method of claim 1, further comprising**

**6[a]: copying a timed portion of the video data to a clip list based on a selection of an item, service, or song of interest by the user.**

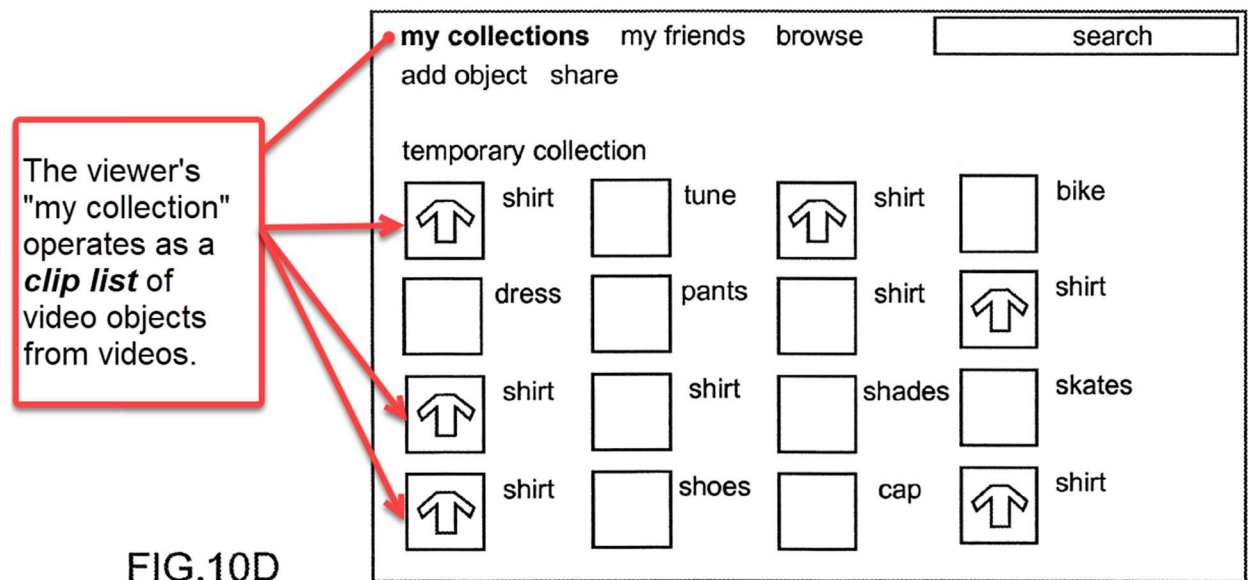
193. In my opinion, Ohliger discloses this limitation. Ohliger discloses a viewer compiling a collection of objects that appear in videos (*copying a timed*

*portion of the video data to a clip list*). EX1006, 25:43-59. This collection mechanism enables viewers to curate personalized sets of product discoveries across multiple videos. Viewers *select[] ... item[s] ... of interest* to add to their collection. *Id.* The system preserves these selections for later access and action. The viewer may then share this collection with others. EX1006, 25:43-26:3.

194. Figure 10C shows the overlay with an option for the viewer (*user*) to select an item of interest to add to their “collection” (*clip list*). A POSITA would have recognized that this collection functionality requires storing references to specific temporal locations within videos where selected objects appear. The implementation captures both the object identification and its temporal context within the source video. Through this collection feature, Ohliger demonstrates *copying a timed portion of the video data to a clip list*, where each collection entry preserves the association between the selected item and its location in the video timeline. Therefore, the viewer’s ability to build persistent collections of video objects establishes the claimed functionality of maintaining a *clip list* containing references to specific *timed portion[s] of the video data*.



195. Figure 10D shows a viewer's "collection" of selected video objects from videos.



196. Ohliger's "collection" identifies video objects (***copying***) with tracked "temporal coordinates" (***timed portion***) within the video (***video data***). EX1006,

11:21-26, 29:66-30:6. The temporal coordinate tracking ensures that collected objects maintain their reference to specific moments within the source video.

197. Patent Owner's Infringement Contentions contend this limitation is met by "allow[ing] you to bookmark videos containing an item, service, or song of interest by clicking on a ribbon indicator." EX1005, 77. This interpretation aligns with Ohliger's collection functionality. Under Patent Owner's view, Ohliger also discloses this limitation by allowing a user to bookmark a selected object for sharing. EX1006, 3:62-63 ("The viewer also has the ability to add the selected object to a collection of objects."). The user "collection" allows the user to "define clusters of objects in a collection and share the entire collection instead of sharing individual objects." EX1006, 26:7-12. This clustering capability transforms individual object selections into organized groups. Effectively, collections "allow users to create a 'wardrobe collection.'" EX1006, 26:13-14. Through these collection mechanisms, Ohliger demonstrates *copying a timed portion of the video data to a clip list*, according to Patent Owner, where the temporal coordinates preserve the connection between collected items and their video source locations. Therefore, whether viewed as bookmarking or collection-building, Ohliger's system enables users to compile references to specific *timed portion[s] of video data* containing items of interest.

198. And even if this limitation is interpreted to require "copying," Ohliger at least renders this obvious because copying the "video objects" to local memory

would give users access to their “collection of video objects” even when their devices are not connected to the Internet. A POSITA would have recognized the technical benefits of local storage for collected items. Offline access represents a standard design consideration for mobile applications, particularly given the intermittent connectivity inherent in mobile device usage.

**(g) Claim 7**

**(i) The method of claim 1, further comprising**

**7[a]: collecting data based on actions taken by the user.**

199. In my opinion, Ohliger discloses this limitation. Ohliger discloses *collecting data based on actions taken by the user* through the “creation of user profiles, which entail the history interaction and sharing of video objects.” EX1006, 11:1-8, 5:44-48, 14:12-21. This data collection encompasses multiple user behaviors within the interactive video system.

200. Ohliger then discloses using the data in the “viewer profile” as part of targeting advertising to each particular viewer for the hyperlinked content associated with each video object. *Id.*, 14:4-21, 14:60-63 (“the VOHE 300 may provide 3rd parties such as retailers/brands/affiliates/advertisers 122 with the profile of an actual video viewer in real-time before making the bid”). The system leverages collected behavioral data to all auction bidders to understand the viewers of the video objects. Therefore, Ohliger’s disclosure of creating viewer profiles from interaction histories

and using this data for targeted advertising establishes the *collecting data based on actions taken by the user* within the interactive video platform.

**(h) Claim 8**

**(i) The method of claim 1, further comprising:**

**8[a]: removing the product information from the one or more of the plurality of visual or audio objects; and associating different product information with the one or more of the plurality of visual or audio objects.**

201. In my opinion, Ohliger discloses this limitation. Ohliger discloses auctioning “the right to assign a hyperlink to a video object” as described for Figure 1’s step 740. EX1006, Fig. 1, 12:29-37. This auction mechanism enables dynamic reassignment of the hyperlink associated with each object. Because the auction allows a different retailer to subsequently change an assigned hyperlink, the *product information* for a video object is *removed* and replaced with *different product information*. EX1006, 13:44-50 (“[A]uction may be triggered by expiration of a retailers/brands/affiliates/advertisers right to associate advertising with a given video object with a hyperlink.”).

202. In my opinion, a POSITA would have recognized that this expiration-triggered auction system necessitates the removal of existing product associations before establishing new ones. Therefore, through the auction-based reassignment of hyperlinks and associated commercial content, Ohliger’s system enables *remov[ing]*

existing ***product information*** and replacing it with ***different product information*** as auction winners change over time.

**(i) Claim 9**

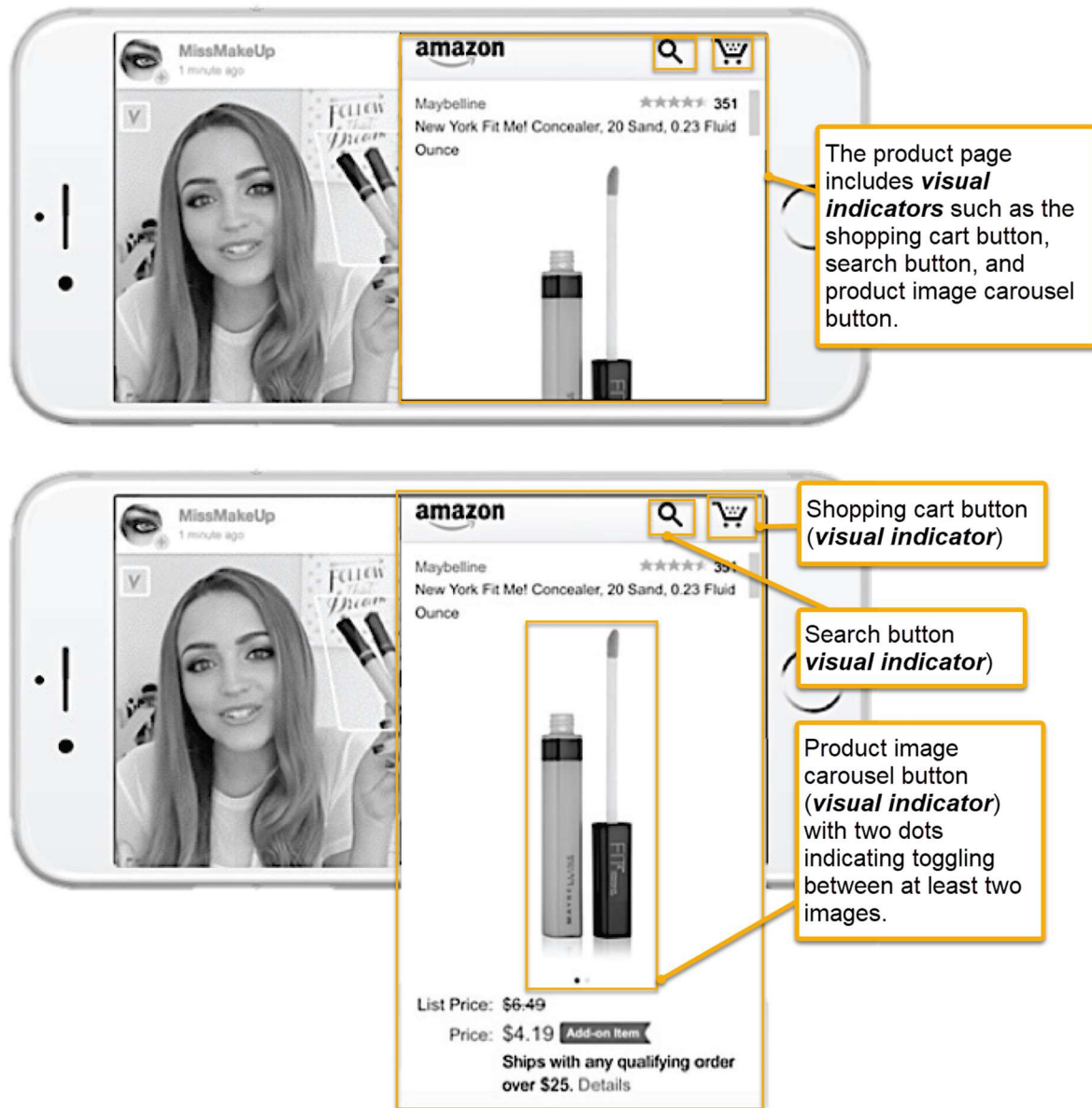
**(i) The method of claim 1, further comprising**

**9[a]: providing a third interactive video layer having one or more visual indicators to the server for distribution with the video media to the user.**

203. In my opinion, Ohliger discloses this limitation. Ohliger discloses a “product pages via a split screen” (***third interactive video layer***). Figure 28 shows an exemplary “product page[]” (***visual indicator[]***): “FIG. 28: No re-direct is required in VeePiO. Users can view product pages via a split screen, giving them a way to browse that product while they watch it in action.” EX1006, 29:19-21.

204. The product page overlay creates a distinct layer separate from both the object overlays and social commenting functionality. A POSITA would have recognized that this third layer provides expanded commercial information beyond the basic overlays of the first layer. Therefore, through the product page overlay, Ohliger discloses a ***third interactive video layer*** containing ***visual indicator[s]*** that enable detailed product browsing concurrent with video playback.

Fig. 28  
Buy Via Split Screen



205. Figure 28 shows the product page for an item appearing in the video based on Amazon.com. The Amazon.com interface demonstrates integration with established e-commerce platforms. A POSITA would have recognized that

leveraging existing commercial infrastructure provides proven transaction processing and fulfillment capabilities. Figure 2 displays *distribution* of the product page to the viewer through the layer server as discussed regarding limitation 1[c][i] above. EX1006, 22:7-11. The server's distribution ensures consistent delivery of the product page layer across multiple viewing devices. Through this architecture, Ohliger establishes that the *third interactive video layer* containing product page *visual indicators* is generated at the server and distributed to viewers alongside the video content and other interactive layers.

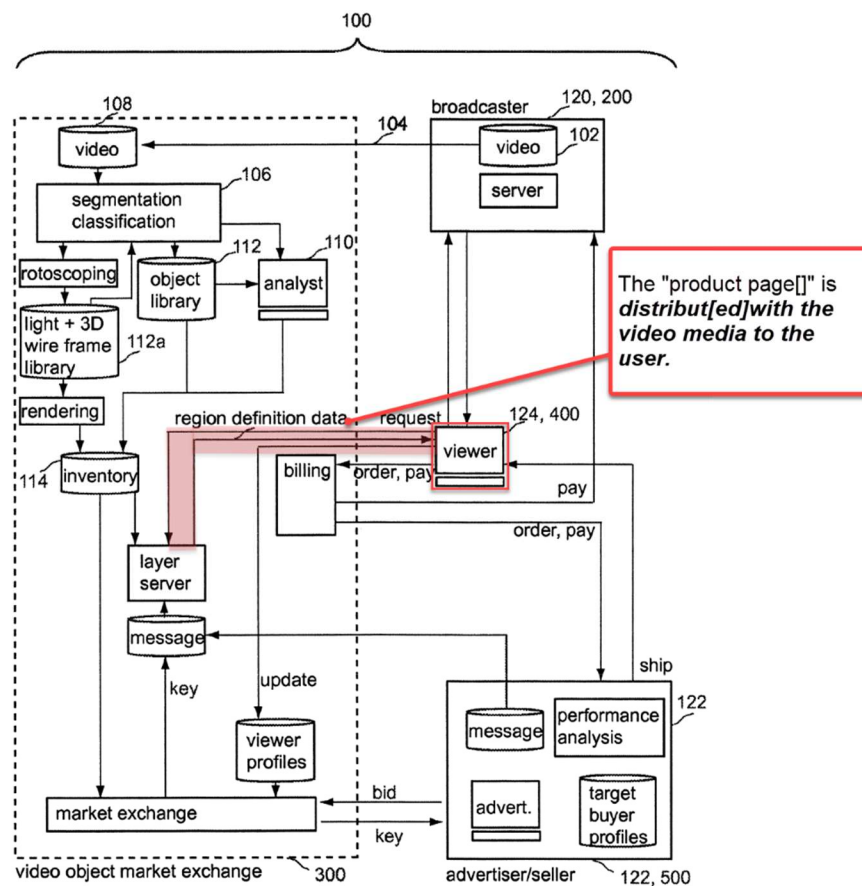


FIG. 2

**(j) Claim 10**

**(i) The method of claim 9,**

**10[a]: wherein the one or more visual indicators include a link for one or more of a second video media, a side room digital layer, a promotional video, a game, and/or merchandise.**

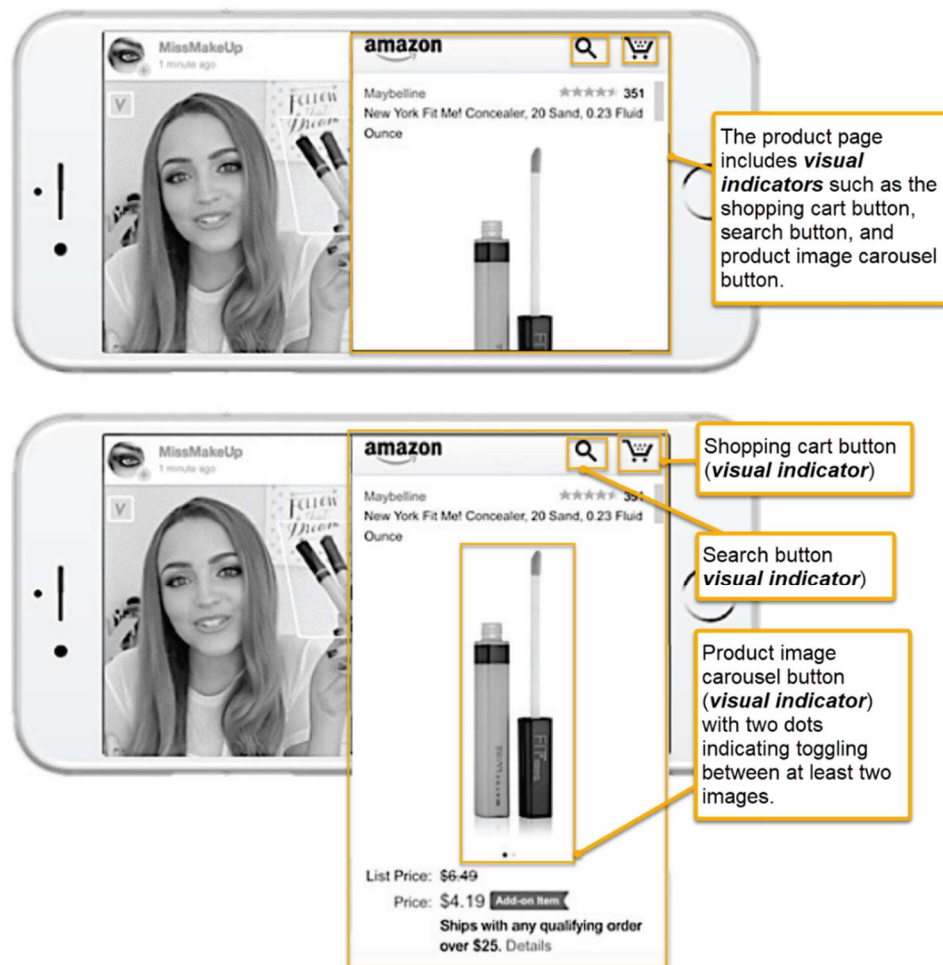
206. In my opinion, Ohliger discloses this limitation. Ohliger discloses a “product pages via a split screen” with *visual indicators includ[ing] a link for one or more ... merchandise*. EX1006, 29:19-21 (“FIG. 28: No re-direct is required in VeePiO. Users can view product pages via a split screen, giving them a way to browse that product while they watch it in action.”).

207. Figure 28’s “product page[]” (*visual indicator[]*) includes interactive buttons such as the search button, the shopping cart button, and the product image carousel button, each of which *link to merchandise* through user selection. These interface elements provide distinct pathways to commercial functionality. A POSITA would have recognized that the shopping cart button enables purchase transactions, the search button facilitates product discovery, and the carousel displays alternative product views or related items. EX1006, 25:17-42.

208. Additionally, Figure 28’s title “Buy Via Split Screen” indicates that a purchase option is available through the “product page[]” overlay. The explicit reference to buying confirms the e-commerce nature of these visual indicators. Through these multiple interactive elements within the split-screen product page,

Ohliger demonstrates that the *third interactive video layer* contains *visual indicators* *includ[ing] a link for one or more ... merchandise*, enabling viewers to browse and purchase products while maintaining visual context of the items in use.

Fig. 28  
Buy Via Split Screen



(k) Claim 11

(i) The method of claim 1,

11[a]: wherein the video data includes a third interactive video layer, wherein the third

**interactive video layer comprises one or more visual indicators.**

209. In my opinion, Ohliger discloses this limitation. See previous discussion regarding Claims 9 and 10, above.

**(I) Claim 12**

**(i) The method of claim 1,**

**12[a]: wherein the video media includes one of a television episode or a movie.**

210. In my opinion, Ohliger discloses this limitation. Ohliger discloses that the video objects appear in video from “television systems” and “[m]otion pictures on film,” *i.e. a television episode or movie*. EX1006, 1:37-47, 13:51-54 (“In one embodiment retailers/brands/affiliates/advertisers can upload a database of images, video objects or 2D/3D wire frames to find matches in *movies* across multiple databases”); Fig. 11A (“show video segments with video object in *movies*”); 41:46-47 (“3603: An application presenting the media. It can be a mobile, web or *television set top box* application”).

211. Ohliger’s disclosure of the system’s capability to process professional entertainment content demonstrates its application to commercial media. A POSITA would have recognized that television episodes and movies represent prime content for interactive shopping experiences, as viewers develop interest in products worn or used by characters.

212. Ohliger also expressly references video objects from the television series “sex and the city” as part of the “object inventory database” in Figure 7. EX1006, Fig. 7 (“HBO comedy-drama, **sex and the city**, cable television series, picture format 480i SDTV etc.”), 7:47 (“FIG. 7 is an exemplary object inventory database”). This reference indicates that the video in question was from that particular television series. The specific identification of this HBO series confirms that Ohliger’s system operates on mainstream television content. Through these disclosures of movies and television series as source content, Ohliger establishes that the interactive video system processes *a television episode or movie*, enabling viewers to interact with and purchase items appearing in professional entertainment media.

| unique object ID | semantic description of object | unique object occurrence ID and reference data | film data                                                                                  | context and verbatim of occurrence                                                                                                                 | metadata viewer profile                                                       | bid per viewer profile channel 1 | bid per viewer profile channel n | winning bid, owner data, run time           |
|------------------|--------------------------------|------------------------------------------------|--------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------|----------------------------------|----------------------------------|---------------------------------------------|
| 1.1.1.1.1.1      | silver necklace                | 1.1.1.1.1.A, x,y,z,t                           | HBO, comedy-drama, sex and the city, cable television series, picture format 480i SDTV etc | Charlotte York (Kristin Davis) having lunch with Miranda Hobbes (Cynthia Nixon) at Sushi Samba discussing birthday present. "...bla, bla, bla...." | target: female, 30y, income 80k, etc, transaction history, viewer profile etc | MMXV: \$1<br>MMXZ: \$3           | MMXV: \$1.5<br>MMXZ: \$2         | Tiffany, December 1, 2008-december 31, 2008 |
| 1.1.1.1.1.1      | silver necklace                | 1.1.1.1.1.B, x,y,z,t                           |                                                                                            |                                                                                                                                                    |                                                                               |                                  |                                  |                                             |
| 1.1.1.1.1.1      | silver necklace                | 1.1.1.1.1.C, x,y,z,t                           |                                                                                            |                                                                                                                                                    |                                                                               |                                  |                                  |                                             |
| 1.1.1.1.1.1      | silver necklace                | 1.1.1.1.1.D, x,y,z,t                           |                                                                                            |                                                                                                                                                    |                                                                               |                                  |                                  |                                             |
| 1.1.1.1.1.1      | silver necklace                | 1.1.1.1.1.E, x,y,z,t                           |                                                                                            |                                                                                                                                                    |                                                                               |                                  |                                  |                                             |
| 1.1.1.1.1.2      |                                |                                                |                                                                                            |                                                                                                                                                    |                                                                               |                                  |                                  |                                             |
|                  |                                |                                                |                                                                                            |                                                                                                                                                    |                                                                               |                                  |                                  |                                             |
|                  |                                |                                                |                                                                                            |                                                                                                                                                    |                                                                               |                                  |                                  |                                             |

**(m) Claim 13**

**(i) The method of claim 1,**

**13[a]: wherein the identifying one or more of the plurality of visual or audio objects is performed manually through input of spatial and/or temporal data for each of the one or more items, services, or songs of interest within the video media.**

213. In my opinion, Ohliger discloses this limitation. As discussed regarding limitation 1[b][ii], Ohliger discloses that the creator of the video can *manually identify specific visual objects* for overlaying. EX1006, 3:9-21 (“identifying objects in video ... manually by using information provided by the creator of the video”). This includes identifying the boundary of the object, *i.e. spatial data of the item of interest*. *Id.*, 12:11-17 (“[A]n analyst may manually identify/configure the object boundaries”), 15:62-65, 18:25-31, 18:31-35, 18:36-38. The manual configuration enables precise delineation of object regions regardless of their shape or complexity.

214. Additionally, Ohliger describes a manual process for “train[ing] the segmentation application” that involves “tracing the outline of a video object.” *Id.*, 13:17-21. This tracing operation captures the *spatial data* by recording the path coordinates that encompass the object. Through these manual identification and boundary configuration mechanisms, Ohliger demonstrates that creators can *manually identify specific visual objects* by defining their *spatial data* within the video frames.

**(n) Claim 14**

**(i) The method of claim 1,**

**14[a]: wherein identifying one or more of the plurality of visual or audio objects is performed automatically by a computer program for visual and/or audio recognition and mapping.**

215. In my opinion, Ohliger discloses this limitation. As discussed above for limitation 1[b][ii], the video segmentation process is a *computer program* of Ohliger that *automatically identifies video objects*. EX1006, 15:6-8 (“[A]ctions which take place without human intervention”), 15:11-12 (“The VOHE 300 can automatically segments video files and automatically classifies the video objects”), 15:20-22 (“Thus, the VOHE 300 may according to some embodiments be a fully automated system.”). The system employs algorithmic processing to detect and classify objects within video frames without requiring manual input.

216. As further discussed, this automated process recognizes visual objects and maps them to product information. The *computer program* analyzes video content using pattern recognition and object detection algorithms to identify items of interest. EX1006, 11:32-37 (“An object of the present invention is to take conventional video content analysis technology ( computer vision algorithms, object tracking algorithms) ... and apply it to the field of marketing communication, advertising and commerce transaction.”), 13:17-21 (“the system may present patterns of content objects it found by statistical pattern recognition.”), 32:30-32

(“After the video has been segmented into frames, an object recognition algorithm can track and identify every frame where a specific object appears.”). A POSITA would have recognized that this automation requires sophisticated image processing capabilities implemented in software such as the video content analysis application 106. EX1006, 16:19-31. Therefore, through the video segmentation process and associated analysis applications, Ohliger discloses a ***computer program*** that ***automatically identifies video objects***, transforming raw video content into interactive media with identified objects ready for overlay generation.

**(o) Claim 15**

**(i) The method of claim 1,**

**15[a]: wherein the first and second interactive layers are unaffected by ad blocking software.**

217. In my opinion, Ohliger discloses this limitation. Ohliger discloses that the ***first and second interactive layers*** include “original, organic” content integrated into the video content. EX1006, 22:42-48; 14:51-59. In my opinion, “original” and “organic” content (*e.g.*, content integrated and core to the viewer application) is ***unaffected by ad blocking software*** because such software is known to block advertising content, *e.g.*, web browser pop-ups to third-party websites, not organic content.

218. In addition, in my opinion, it would have been obvious to a POSITA that the “original, organic” content of the ***first and second interactive layers*** would

be *unaffected by ad blocking software* because that was a well-known benefit to organic content at the time. EX1009, ¶0019. The distinction between organic and third-party content was widely understood. It was well-known that consumers used advertisement blocking software, and it operated by preventing the consumer software from receiving web requests “not associated with currently presenting media and/or requests that are associated with known advertising servers.” EX1009, [0018].

219. A POSITA would understand that the advertising content, *e.g.* hyperlinks, in Ohliger would be embedded in the overlays with the organic content, *e.g.* creator descriptions, product information, and viewer comments, and further incorporated with the video content to circumvent the ad-blocking software. EX1009, [0019]. This integration strategy leverages the native content delivery mechanism.

220. Additionally, Ohliger’s system includes a non-web browser software application for viewers that would be known to be *unaffected by ad blocking software*. EX1006, 4:1-3, Figure 13. A POSITA would have understood that consumer ad-blocking software included browser extensions that operated only on content within web browsers such as Internet Explorer, Safari, Google Chrome, and Firefox. EX1019, 1:38-49. Native applications bypass these browser-based blocking mechanisms entirely. In addition, as discussed in limitation 1[c], it would have been

obvious to integrate the overlays into the video, which has the benefit of being *unaffected by ad blocking software*. Therefore, through both the organic nature of the content and the native application delivery, Ohliger's *first and second interactive layers* would be *unaffected by ad blocking software*.

(p) **Claim 16**

(i) **The method of claim 1,**

**16[a][i]: wherein the interaction and the past interaction activity comprise a public or private conversation between the user and the at least one guest,**

221. In my opinion, Ohliger discloses this limitation. As discussed regarding limitations 1[c][iii] and 1[c][iv], Ohliger discloses an interactive comments layer with *present and past interactions* that are *public*.

(ii) **16[a][ii]: wherein the plurality of different conversations comprises at least a public conversation which enables all viewers of the video media to join and a private conversation which enables a permitted subset of all viewers of the video media to join.**

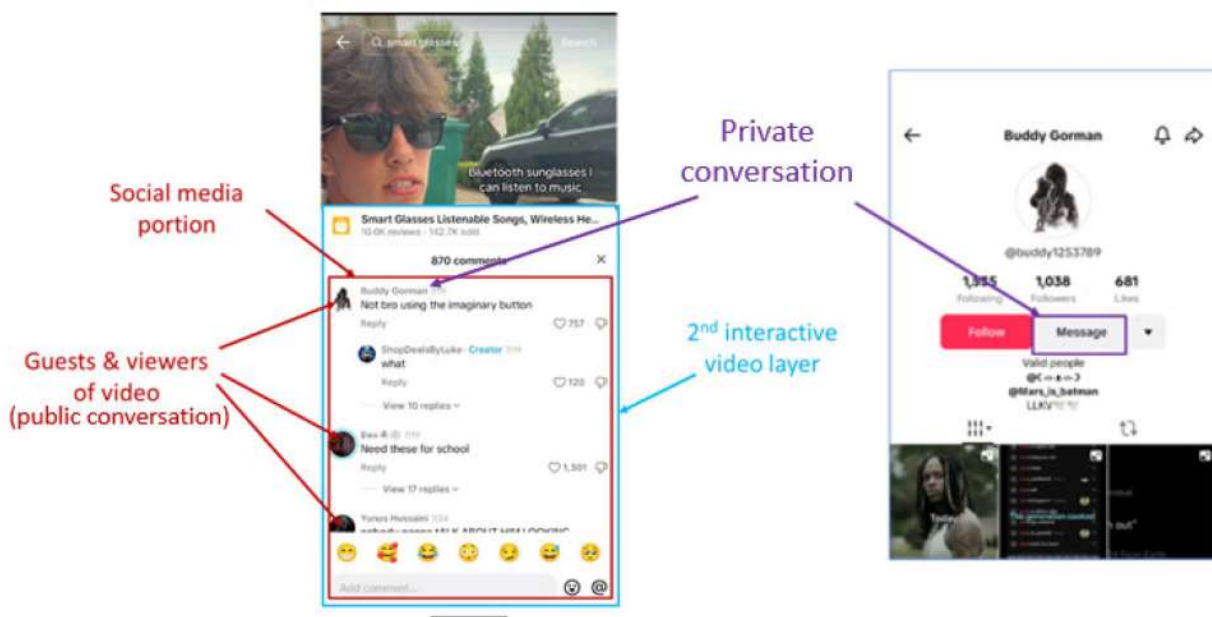
222. In my opinion, Ohliger discloses this limitation. As discussed above in limitation 1[c][iii] and 1[c][iv], Ohliger discloses that comments include *public conversation[s]* from *all viewers*. The sub-comment threads associated with each object create open forums where any viewer can participate. Ohliger discloses this public accessibility enables community-wide discussions around specific items appearing in the video content. EX1006, 28:42-45 ("VeePiO thrives on community.

IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

Commenting is easy and encouraged via standard pathways like the feed, and through product sub-comments. Beyond that VeePiO media can be shared anywhere online”).

223. Ohliger also discloses *a private conversation between a permitted subset of viewers* under PO’s interpretation. PO’s Infringement Contentions accuse private messaging functionality that is not overlaid onto the video. EX1005, 111-112. This interpretation contends that private conversations need not appear as overlays to satisfy the limitation.



224. Under PO’s view, Ohliger discloses in Figure 15 a button to “[s]end a message” to a specific user, which is a *private conversation* between a *permitted subset of all viewers*. This messaging functionality enables direct user-to-user communication outside the public comment threads. A POSITA would have

recognized that the “[s]end a message” button initiates a restricted communication channel accessible only to the sender and recipient.

225. The implementation creates a two-party subset from the broader viewer population, establishing the *permitted subset* through the messaging relationship. Unlike the public sub-comments visible to all viewers, these private messages remain confined to the participating users. Therefore, Figure 15’s messaging capability demonstrates Ohliger’s support for *private conversation[s]* between a *permitted subset of all viewers*, complementing the public discussion forums with targeted interpersonal communication channels.



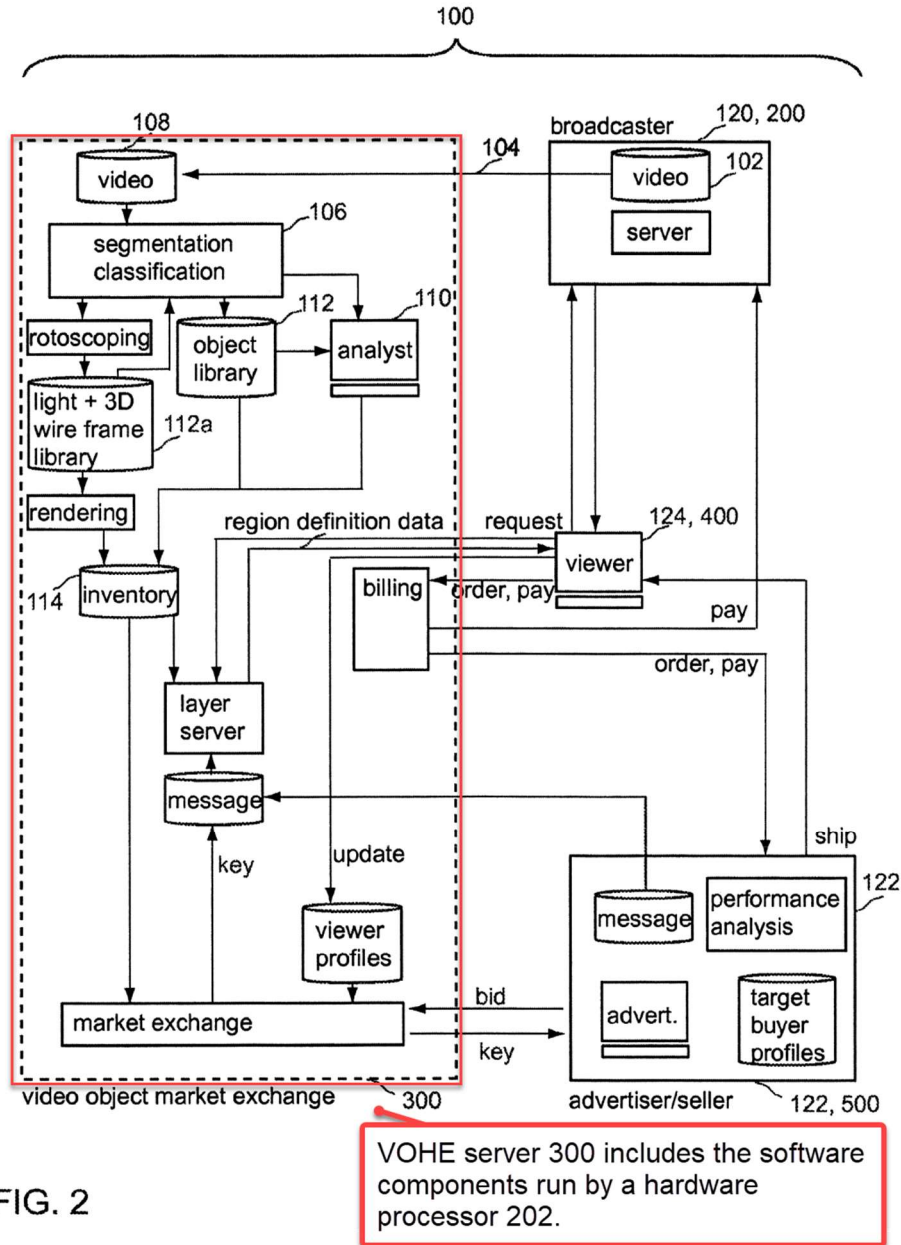
(q) Claim 17

(i) 17[pre]: A non-transitory computer readable media encoded with instructions that, when

**executed by a processor, cause the processor to carry out at least the following steps:**

226. In my opinion, Ohliger discloses this limitation. Ohliger discloses “VOHE server 300” which “includes a processor 202 which is connected via BUS 204 to a mass storage device 206, Read-Only-Memory (ROM) 208 and Random Access Memory 210 (which may be volatile or nonvolatile).” EX1006, 16:10-13. These hardware components establish the physical computing infrastructure necessary for executing the claimed method. This server performs the corresponding “steps” of claim 17. EX1006, 19:13-18, 22:4-7, 26:32-46.

227. Figure 2 provides a diagram that includes VOHE server 300’s software components (*non-transitory computer readable media encoded with instructions*). EX1006, 15:27-28, 16:28-31, 17:27-29, 19:24-26, 26:52-55. The software architecture comprises multiple applications. These applications direct the processor 202 to perform the video processing, overlay generation, and distribution operations described throughout Ohliger. Therefore, VOHE server 300 includes the *non-transitory computer readable media encoded with instructions that, when executed by a processor, cause the processor* to perform the claimed method steps.



228. Thus, while VOHE server 300 may include multiple software components, as shown in Figure 2, that software is all executed by the same hardware processor 202. EX1006, 16:10 (“Server 300 includes a processor 202”), 19:14-18 (“The system includes a server such as server 300 including a computer having a processor, random access memory, read-only memory, and mass storage

memory, the server includes one or more video files stored in a video database.”).

Accordingly, the same processor performs all steps of claim 17 addressed herein.

229. Moreover, PO disputes that the claims require the same processor to perform all steps recited in claim 17. Under this broader interpretation, a larger number of system architectures and their implementations would be within scope of the claims. This expansive reading accommodates both a single processor and distributed processing implementations. To the extent more than one processor is allowed to perform the steps of claim 17, there should be no dispute that Ohliger’s software-implemented system is performed by one or more processors. EX1006, 46:12-20.

**(ii) 17[a]-17[b]**

230. In my opinion, Ohliger discloses this limitation. *See* limitation 1[b]-1[b][iv] and 1[c]-1[c][vi].

**(r) Claim 18**

- (i) 18[pre]: A computer-implemented system comprising: a digital processing device comprising: at least one processor, an operating system configured to perform executable instructions, a memory, and a computer program including instructions executable by the digital processing device to create an application for providing information on items,**

**services, or songs in a video media, the  
application comprising:**

231. In my opinion, Ohliger discloses this limitation. Claim 18's preamble is oddly written in that it requires a *digital processing device*, but that device does not actually include the *software module[s]* recited in the body of the claim. Rather, the *digital processing device* must only *create an application*, and it is this *application* that *provid[es] information on items, services, or songs in a video media* and contains the *software module[s]* recited in the body of the claim. This unusual claim structure requires careful analysis of how Ohliger's disclosures map to the claimed elements. Nonetheless, Ohliger discloses or renders obvious the preamble of claim 18.

232. Ohliger's system 100 is *[a] computer implemented system*. EX1006, 8:58-9:23, 15:27-44, 19:13-37. The computational architecture provides the necessary processing infrastructure. System 100 includes VOHE server 300 (*a digital processing device*). EX1006, Fig. 2, 16:10-18, 19:13-37. Ohliger discloses VOHE server 300 *comprising: at least one processor, an operating system configured to perform executable instructions, [and] a memory*. EX1006, Fig. 2 (showing embodiment of VOHE server 300), 16:10-13, 19:14-18, 46:3-11. These hardware and software components establish the platform for application execution.

233. Ohliger discloses *creat[ing] an application* through compilation or intermediate code. EX1006, 46:6-11. This creation process transforms source code

into executable applications. As discussed above for claim 1, Ohliger discloses *application[s] for providing information on items, services, or songs in a video media*. EX1006, 13:22-40, 19:24-37, 26:31-27:22. These *application[s]* perform the steps of claim 1, addressed above, and contain the *software module[s]* of claim 18, addressed below. EX1006, 8:58-9:15. Therefore, through VOHE server 300's hardware infrastructure and application creation capabilities, Ohliger comprehensively discloses the *digital processing device* that *creat[es] an application for providing information on items, services, or songs in a video media*.

234. According to PO, the “digital processing device” in the preamble “provides definite structure for the software modules.” EX1017, 18. Even if that means the claimed *digital processing device* must then execute the claimed *application*, it would be at most obvious for VOHE server 300 to compile the application(s) executed thereon. The server's dual role as both compiler and executor represents efficient system design.

235. VOHE server 300 (alone or with integrated broadcast server 200 functionality) is the only hardware disclosed in Ohliger as the platform for hosting VeePIO, and a POSITA would thus understand that both the compiling and the running of the application disclosed in Ohliger occurs on VOHE server 300. The singular hardware platform necessitates this consolidated functionality. In addition, it would be obvious to use the same server for both purposes, since VOHE server

300 has the necessary hardware for compiling its applications, and compiling applications on the same VOHE server 300 used to run them avoids the need to transfer those applications post-compilation to a different server and the associated risk of incompatibility.

236. A POSITA would have recognized that local compilation eliminates network transfer delays and ensures binary compatibility with the target execution environment. Thus, Ohliger discloses or renders obvious all structure PO attributes to the claimed “software module[s].”

237. PO disputes that the claims require the same application to include all software modules recited in claim 18. This interpretation broadens the claim scope considerably. To the extent more than one applications are allowed to contain the software modules of claim 18, in my opinion, there should be no dispute that Ohliger’s software-implemented system is performed by one or more applications. The modular software architecture disclosed in Ohliger accommodates multiple application configurations while maintaining functional cohesion.

238. In my opinion, even if claim 18 requires all *software module[s]* to be part of the same *application*, Ohliger renders this obvious. The consolidation of modules represents a straightforward design choice. All applications at issue reside on the same VOHE server 300. Thus, dividing or combining functionality into one or more applications is a simple design choice. A POSITA would have recognized

that module organization within applications involves routine engineering decisions.

Since the functionalities at issue here are all intended to be used together, combining those functionalities into one application would simplify design by limiting the need for coordination between applications for purposes of timing and memory access.

EX1006, 16:28-31, 46:21-28.

239. Finally, Ohliger discloses performing the steps of claim 18 in the order recited in the claims. In my opinion, to the extent the order of steps for *processing* is required by this limitation, Ohliger meets this ordering. The temporal sequence ensures proper data flow through the system. For example, Figure 1 provides the following ordered steps:

240. (1) limitation 18[a]’s *receiving a request* with creator’s upload prior to step 700’s “[s]egment[ing] video into frames,”

241. (2) limitation 18[b]’s *linking the items* with step 715’s “[c]reator ... add[ing] comments, pictures and/or links,”

242. (3) limitation 18[c]’s *generation of a first interactive layer* and then 18[d]’s *generation of a second interactive layer* with step 745’s “[l]ink overlay box,” and

243. (4) limitation 18[e]’s *transmitting* with step 750’s “[s]erving object information.” EX1006, Fig. 1.

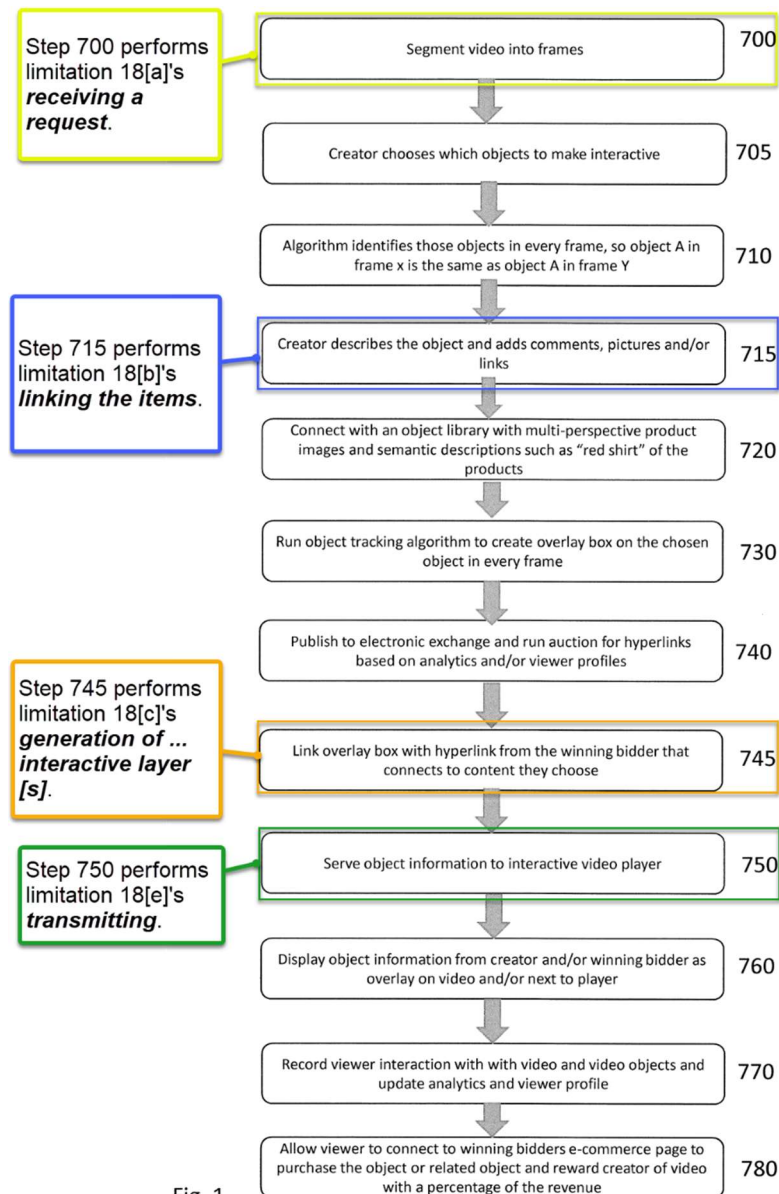


Fig. 1

244. In addition, in my opinion, Ohliger discloses or at least renders obvious that the steps identified for limitations 18[a]-[e] below can be performed in the order recited in the claims. EX1006, 13:6-11. For example, sub-comments overlaid for a particular video objects are specific to that video object. Ohliger, 11:11-19. A POSITA would understand that the overlay for sub-comments for the specific video object (*second interactive video layer*) is generated only after first creating and

linking overlays with the visual outlines corresponding to that specific video object (*first interactive video layer*). The dependency between these layers supports this sequence. The second layer requires the first layer's object identification and boundary definitions to properly associate comments with their corresponding visual elements.

245. At a minimum, it would be obvious for the system in Ohliger to generate the sub-comments overlay specific to the video object later in the sequence because doing so is more efficient than the alternative: first generating a generic overlay, then creating and linking the video object, and then backtracking to populate the sub-comments overlay with comments, such as those from the creator, applicable to that specific video objects. A POSITA would have recognized that this forward-sequential processing eliminates redundant operations and reduces computational overhead. The ordered approach ensures that each layer builds upon the previous layer's established data structures. Therefore, both Ohliger's disclosed system architecture and obvious implementation efficiencies support performing the claimed steps in their recited order.

- (ii) **18[a]: a) a software module for receiving a request for the generation of information regarding items, services, or songs presented in**

**a video media from a first processing device of a user;**

246. In my opinion, Ohliger discloses this limitation. When a creator chooses a specific video to make interactive, the creator *request[s] for the generation of information regarding items, services, or songs presented in a video media.*

247. Ohliger discloses a video creator selecting a video from their device (*first processing device of a user*). EX1006, 28:57-60, Fig. 20. The creator's mobile device initiates the interactive video creation process. After selecting the video, Ohliger discloses the video creator "adding interactivity" to the video through the application on their mobile device. EX1006, 28:61-67, 29:1-4, Fig. 21.

248. To add interactivity to a video, the creator, through their mobile device and application, *requests* step 700's processing of the video by VOHE server 300. EX1006, Fig. 2, 28:57-67, 34:36-44. The request triggers server-side analysis of the video content. Step 700 "yields a list of video objects throughout the video (including the temporal coordinates and region definition data for each object)." EX1006, 11:54-57. A POSITA would have recognized that this object list constitutes the requested information about items appearing in the video, including temporal and spatial coordinates of each object.

249. Therefore, when a creator selects a video to make interactive, it *requests* that step 700 generate the "temporal coordinates and region definition data

for each object” that allows the “[c]reator [to] choose[] which objects to make interactive” in step 705. EX1006, Fig. 1. The creator’s initial video selection and request for interactivity initiates the generation of comprehensive object information necessary for subsequent overlay creation. Through this request-triggered processing sequence, Ohliger demonstrates *receiving a request for the generation of information regarding items, services, or songs presented in a video media from a first processing device of a user.*

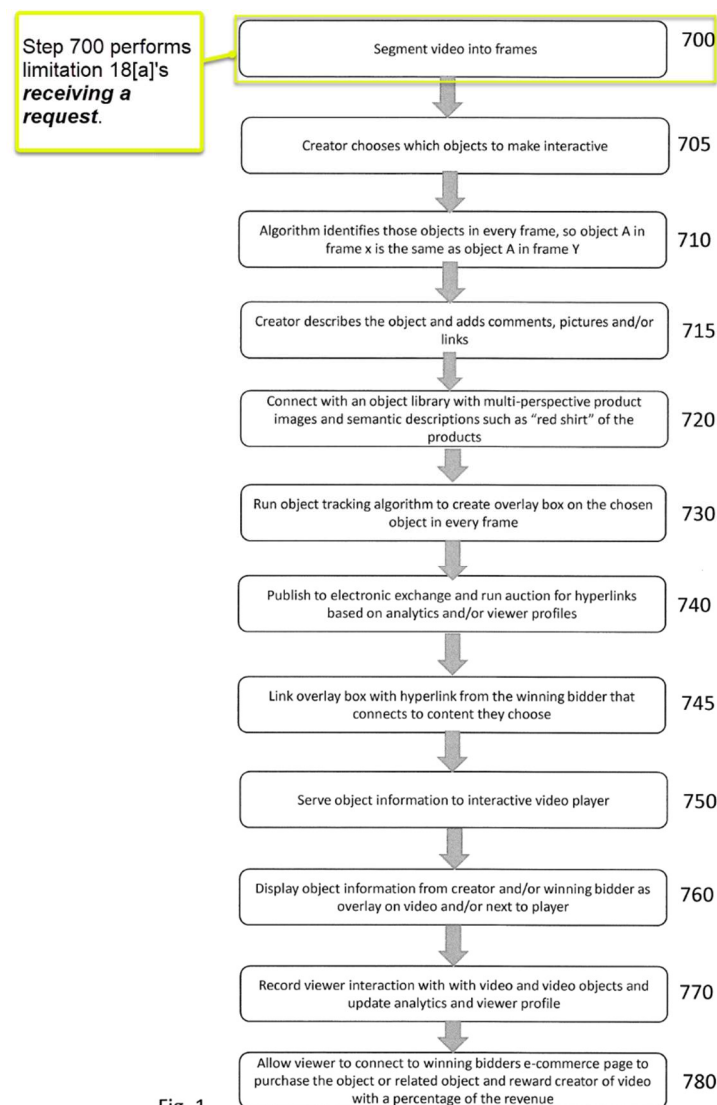


Fig. 1

(iii) **18[b]: b) a software module for linking the items, services, or songs with marketing and/or purchasing information;**

250. In my opinion, Ohliger discloses this limitation. *See* limitation 1[b][i] and 1[b][ii].

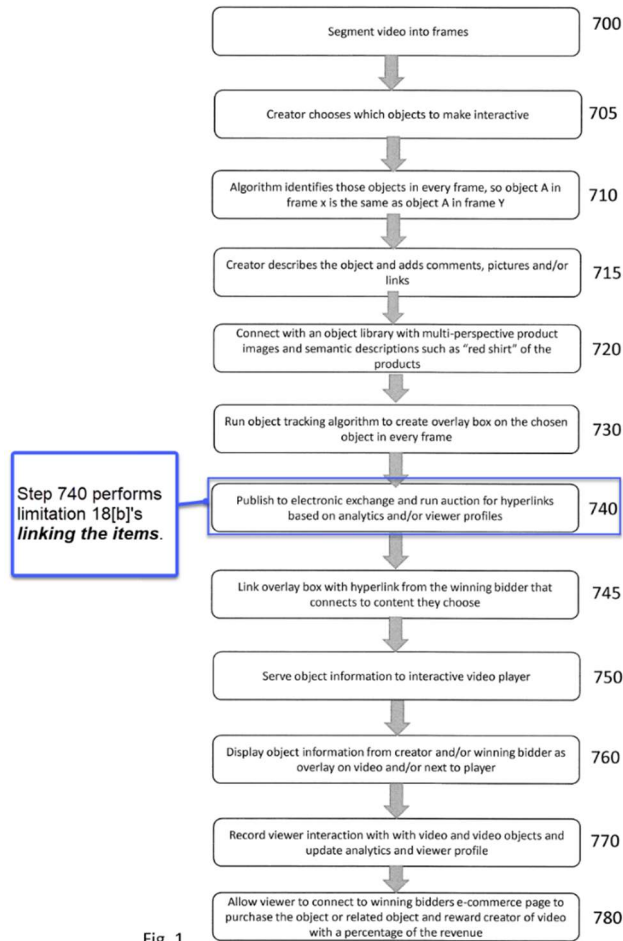
251. In my opinion, to the extent *linking* is different from the *associating* in limitations 1[b][ii], Ohliger discloses *linking the items ... with marketing and/or purchasing information* through the hyperlink auction process. The linking mechanism creates direct connections between identified objects and the website chosen by the winning bidder. EX1006, 12:29-37, 14:25-30, 21:45-62.

252. Figure 1's step 740 describes an "automated auction process" whereby "retailers, brands and affiliates/advertisers ... bid for the right to assign a hyperlink to a video object that takes a viewer of the video who has interacted with that object to a website where that object or any object the winning bidder chooses can be purchased, or to any other content the winning bidder would like to associate with that object." EX1006, 12:29-37. This auction-based assignment establishes the commercial linkage.

253. Ohliger *link[s] the items* through hyperlinks to the winning bidder's website (*marketing and/or purchasing information*). A POSITA would have recognized that these hyperlinks provide the technical mechanism connecting visual objects to their corresponding commercial destinations. The implementation creates

navigable pathways from object selection to product information and purchase options.

254. Ohliger discloses that the VOHE 300 performs step 740 through an “automated bidding application” (*a software module*). EX1006, 22:19-27, 22:42-51. This software module manages the auction process and establishes the hyperlink associations based on winning bids. Therefore, through the automated bidding application’s creation of hyperlinks connecting video objects to commercial websites, Ohliger discloses *a software module configured to link the items ... with marketing and/or purchasing information*.



- (iv) **18[c]: c) a software module for the generation of a first interactive layer to display with the video media, wherein said first interactive layer contains the items, services, or songs of interest linked to the marketing and/or purchasing information;**

255. In my opinion, Ohliger discloses this limitation. *See* limitation 1[b][iv].

- (v) **18[d][i]: d) a software module for the generation of a second interactive layer to display with the video media and optionally the first interactive layer,**

256. In my opinion, Ohliger discloses this limitation. *See* limitation 1[c]-1[c][vi].

257. Ohliger discloses the video and comments are *optionally* displayed with the video object animation, outline, or halo (*first interactive layer*). Specifically, Ohliger discloses that video object animations, outlines, or halos can be turned “off by tapping the ‘V’ watermark a third time. (3rd Screen).” EX1006, 29:5-10. This toggle functionality provides viewer control over overlay visibility.

258. There is no indication that the “V” button functionality is not enabled when viewing the comments overlay (*second interactive layer*) such as in Figure 26. Nor is there any suggestion that the “V” button would affect the display of the comments overlay (*second interactive layer*). A POSITA would have recognized that these layers operate independently, with separate display controls. Therefore, the “V” button’s toggling of the outlines or halos around items allows the viewer to *display ... optionally the first interactive layer* with the comments overlay.

259. In addition, in my opinion, it would have been obvious to a POSITA to enable viewers to *optionally* display outlines or transparent halos around items appearing in videos while displaying comment overlays on videos. The technical implementation involves maintaining separate display states for each layer. Comment overlays enable viewers to post comments, and posting comments may benefit from providing viewers with clear and unobstructed views of particular items in videos.

260. Therefore, in my opinion, a POSITA would be motivated to enable “V buttons” to grant viewer control to view or hide (*display with the video media ... optionally*) outlines or halos around items in videos (*first interactive layers*) while displaying comment overlays (*second interactive layers*). The user interface benefits from granular control over visual elements. A POSITA would have reasonable expectation of success because Ohliger already discloses “V button” functionality, and it would be predictable to extend it to operate while displaying comment overlays. Through either the disclosed functionality or this obvious extension, Ohliger enables *optionally* displaying the *first interactive layer* concurrent with the *second interactive layer*.

**(vi) 18[d][ii]-18[d][vi], 18[e]**

261. In my opinion, Ohliger discloses this limitation. *See* limitation 1[c]-1[c][vi] and 1[d].

**(s) Claim 19**

**(i) 19[pre]**

262. In my opinion, Ohliger discloses this limitation. *See* Claim 18.

**(ii) 19[a][i] and 19[a][ii]**

263. In my opinion, Ohliger discloses this limitation. *See* limitations 16[a][i] and 16[a][ii].

**B. Ground 2: Ohliger and Sakhartov**

**1. Background**

264. U.S. Pat. No. 8,166,500 (“Sakhartov”) was filed May 22, 2008, published June 18, 2009, and granted April 24, 2012. Sakhartov is prior art under 35 U.S.C. 102(a). Sakhartov discloses “generating interactive content from video content” where “at least one interactive element is integrated onto the video.” EX1013, Abstract.

**2. Proposed Combination**

265. Ohliger’s system would be modified, according to Sakhartov’s teachings, to use Sakhartov’s “merging process” to “aggregate[] and combine[] to form the final video.” EX1013, 7:56-62. As part of this modification, Ohliger’s “overlay” and “video content” would be merged into Sakhartov’s “final video” prior to sending to Ohliger’s video viewer 124.

**3. Motivation to Combine & Reasonable Expectation of Success**

266. In my opinion, A POSITA would have been motivated to apply Sakhartov’s teachings to Ohliger’s video overlay system, and would have had a reasonable expectation of success in combining the references.

267. In my opinion, as a threshold matter, Ohliger, Sakhartov, and the ’316 Patent are all analogous art because the ’316 Patent and the prior art references are all in the same field of endeavor, *i.e.*, overlaying interactive elements on video content. For example, the ’316 Patent relates to “interactive video layers” for

“associating product information to one or more ... objects” within the video which are distributed with and displayed with streaming video content. EX1001, Abstract, 2:22-50. Ohliger and Sakhartov similarly relate to generating interactive content for videos that is displayed as overlays on the video.

268. Ohliger is directed to “identifying objects in ... video” and making them interactive with a “clickable space around the object in ... every frame.” EX1006, Abstract. Sakhartov is directed toward “generating interactive content from video content” by “identifying at least one product within the video content and converting the product into an interactive element within the video.” EX1013, Abstract. These are analogous to the ’316 Patent.

269. In my opinion, POSITA would have looked to Ohliger and Sakhartov when considering how to implement known systems for interactive video content related to items that appear within the video. A POSITA would have further been motivated to combine Ohliger and Sakhartov for the following reasons.

270. *Circumvent ad-blocking software*: In the four corners of Ohliger, it was understood that the goal of Ohliger was to present a unique form of advertising to viewers of video content. EX1006, 14:31-45. This commercial focus drives technical implementation decisions throughout the system. With that in mind, in my opinion, a POSITA would have been motivated to incorporate known ways of further improving the advertising capabilities of the system.

271. A POSITA would have known that the advertising elements of the system provide revenue opportunities for the operator of the system. Indeed, Ohliger is directed to auctioning off object overlays to advertisers. The auction mechanism creates direct financial incentives for maximizing ad delivery effectiveness. Thus, preventing the overlays from being blocked by ad-blocking software would have provided the benefit of increased revenue to advertisers and therefore higher auction prices for video creators.

272. At the time, it was well established that ad-blocking software hindered digital advertising by preventing ad content from being received or displayed by end user devices via blocking network communications that were known to include online advertising content. EX1009, [0018-0019], EX1010, 11:55-59. This widespread blocking created substantial revenue losses across the digital advertising ecosystem.

273. A known technique for circumventing ad-blocking software was merging the organic content, such as a video, with the advertising content so that the advertising content could not be distinguished from the organic content at the network layer. EX1009, [0019], EX1010, 11:55-59. This integration approach defeats pattern-based blocking mechanisms.

274. A POSITA would have understood that the advertising content, *e.g.* hyperlinks, in Ohliger would be embedded in the overlays with the organic content,

e.g. creator descriptions, product information, and viewer comments. Therefore, further incorporating those overlays containing the advertising content into the video content would have circumvented ad-blocking software. EX1009, [0019].

275. In my opinion, a POSITA would have been motivated to integrate the merging process of Sakhartov to circumvent the well-known techniques of ad-blocking software, and, thus, improve the advertising capabilities of the system. The technical implementation would involve server-side rendering of overlays into the video stream, creating an integrated data package resistant to selective blocking. This modification represents a predictable application of known anti-blocking techniques to Ohliger's interactive video platform.

276. ***System Operational Improvements:*** In my opinion, a POSITA would have understood there would also have been secondary benefits to merging the video content and overlays including simplifying the storage, management, and tracking of the merged video data as opposed to maintaining each separately. The consolidation reduces system complexity across multiple operational dimensions.

277. With a single video file containing the video and overlay content, these processes would be more predictable and manageable for system operations because there would have been fewer dynamic components requiring synchronization and communications amongst them in providing the complete video with overlays to the

viewers. A POSITA would have recognized that eliminating multi-component synchronization reduces potential points of failure and timing conflicts.

278. The unified data structure provides additional technical advantages. Storage requirements simplify when managing single files rather than coordinating multiple associated data streams. Tracking viewer interactions becomes more straightforward when all interactive elements exist within the same data package. Content delivery networks can cache and distribute single merged files more efficiently than maintaining temporal synchronization between separate video and overlay streams. A POSITA would have recognized that merging the content, or even a subset of the content, would provide these benefits.

279. ***Design choice:*** Merging video content with interactive layers into single video data files would have been one of a limited number of design choices available to software developers. In my opinion, a POSITA would have recognized that Ohliger's video content and overlays could be merged together or kept separate.

280. As was well-known at the time, digital multimedia formats served as containers encapsulating video and audio content as well as supplemental content such as subtitles, navigation menus, images, hyperlinks, and interactive advertisements. EX1021, [0004], [0020]. These established formats provided proven frameworks for content integration.

281. For example, Apple's QuickTime video files included "sprite tracks" for providing video content interactivity, and Adobe Flash media players allowed for interactive video content as well. EX1020, [0242]-[0245]; [0005]. A POSITA would have recognized these platforms as industry-standard solutions demonstrating the technical feasibility of merged interactive content.

282. A POSITA would have understood that these multimedia formats would have been available for merging Ohliger's overlays with video content into one multimedia file. EX1021, [0023], EX1020, [0005]. The existing infrastructure and widespread support for these formats would have facilitated implementation. QuickTime and Flash players had large installed bases and were generally provided free of charge, ensuring broad compatibility with viewer devices.

283. Therefore, in my opinion, a POSITA would have reasonably expected to succeed in following Sakhartov's teachings for merging video content and overlays into single video data files. The combination of limited design alternatives, established multimedia container formats, and proven interactive video technologies creates a clear path for implementation.

284. *Ohliger's system is suitable because Ohliger describes sending the video content and overlays to the viewer together:* In my opinion, a POSITA would have reasonably expected to succeed in combining Sakhartov's "merging process" for overlays and video content with Ohliger because Ohliger discloses serving

overlays and video content to viewers from single servers. EX1006, 22:7-11. The consolidated server architecture provides an ideal platform for implementing the merging operations.

285. Merging overlays and videos into combined “final videos” would have required minimal changes to that server’s software. A POSITA would have recognized that the existing server infrastructure already handles both data streams, making integration a straightforward modification. The technical implementation would involve adding a rendering step to combine the overlay and video layers before transmission.

286. Further, Ohliger discloses the “video content analysis application” running on the server that could have performed the merging processes to create the final video files. EX1006, 16:19-31. These existing processing capabilities demonstrate the server’s computational resources for handling video manipulation tasks. Therefore, Ohliger’s system architecture and hardware components would have been suitable for combining with Sakhartov. The server’s processor, memory, and storage systems provide adequate resources for the merging operations. A POSITA would have understood that implementing Sakhartov’s merging process requires only leveraging Ohliger’s existing infrastructure with routine software modifications. The combination represents a predictable variation achievable

through known video processing techniques, ensuring a reasonable expectation of success in implementing the merged video data approach.

287. *Software modification would be predictable*: As discussed above, only Ohliger's software would have changed. The hardware infrastructure remains constant while implementing the merging functionality. In my opinion, given the generally-acknowledged predictability of software arts, a POSITA would have had reasonable expectation of success in making proposed enhancements.

288. Software modifications follow well-established patterns in video processing applications. Changes would have been predictable and minimally impactful on overall systems given design modularity and ability to implement changes in specific components for storing and serving video content and overlays. A POSITA would have recognized that modular architectures, like Ohliger's, facilitate isolated updates without disrupting adjacent system components.

289. Software changes would have involved processing video files using well-understood video file formats and associated software libraries. The technical implementation leverages existing tools and frameworks. Industry-standard libraries for video manipulation, provided software libraries to support merging video streams with overlay content. This existing software would have simplified the merging process.

290. Therefore, the proposed enhancement represents routine software development within established technical boundaries. A POSITA would have understood that combining Sakhartov's merging approach with Ohliger's system involves applying known video processing techniques through known and widely available software tools. The modification requires no hardware changes, no novel algorithms, and limited or no departure from standard video processing methodologies. This technical predictability, combined with the modular nature of Ohliger's software architecture, ensures a reasonable expectation of success in implementing the merged video data approach.

#### **4. Analysis**

291. The analysis for Ground 2 incorporates by reference the analysis for Ground 1. Below, Petitioner provides additional analysis for individual claim limitations that further rendered obvious by the proposed combination of Ohliger and Sakhartov (in addition to being rendered obvious by Ohliger alone).

##### **(a) Claim 1**

##### **(i) 1[c]: generating a video data comprising the video media, the first interactive video layer, and a second interactive video layer**

292. In my opinion, the combination of Ohliger and Sakhartov discloses this limitation. Ohliger describes the video (*video media*) and overlays (*the first interactive video layer, and a second interactive video layer*) being distributed to viewers by the broadcaster. EX1006, 22:7-11 ("The overlay is supplied by the

VOHE server 300 to the broadcaster 120 and in turn from the broadcaster 120 to the viewer 124 along with the underlying video.”). This distribution mechanism provides the separate components to the viewing device.

293. Sakhartov discloses merging interactive layers with the “original video data” into a “final interactive video data file” (*generating a video data*). EX1013, 4:44-50, 7:56-62, 10:5-19. The merging process consolidates multiple data streams into a unified file structure. A POSITA would have recognized that Sakhartov’s merging technique transforms Ohliger’s separate distribution model into integrated data delivery.

294. The combination operates by applying Sakhartov’s merging process to Ohliger’s video and overlay components prior to distribution. Ohliger’s VOHE server 300 would execute the merging operation, combining the video media with both interactive layers into a single data file. The resulting merged file contains all interactive elements embedded within the video stream, creating the *video data* that includes *the video media, the first interactive video layer, and a second interactive video layer*. Therefore, through Ohliger’s multi-layer interactive video system enhanced with Sakhartov’s merging technique, the combination discloses *generating a video data*.

**(b) Claim 17**

295. In my opinion, the combination of Ohliger and Sakhartov further renders limitation 17[b] obvious for the reasons set forth above for limitation 1[c].

**C. Ground 3: Ohliger and Kaiserlian**

**1. Background**

296. U.S. Patent No. 8,332,477 (“Kaiserlian”) was filed May 24, 2012 and issued December 11, 2012. Kaiserlian is prior art under 35 U.S.C. 102(a). Kaiserlian describes a method “for presenting related conversation” including “a conversation view ... [with] a flattened representation of the portion of the conversation.” EX1014, Abstract.

**2. Proposed Combination**

297. Ohliger’s system would be enhanced, according to Kaiserlian’s teachings, to use Kaiserlian’s “enhanced conversation view” with “conversation thread[ing]” in Ohliger’s interactive overlay with comments. EX1014, 14:42-47. As part of this modification, Ohliger’s comments for video objects would allow for reply messages that are displayed with “indent levels that indicate the reply depth of each message ... indicat[ing], for example, the number of replies that are in the sequence that connects a message to, for example, a first message in a conversation thread.” EX1014, 14:48-55. Kaiserlian applies reply depth thresholds that omit messages with reply depths that exceed predetermined maximums. EX1014, 15:19-30. Additionally, “the conversation view 450 compresses or collapses the portion

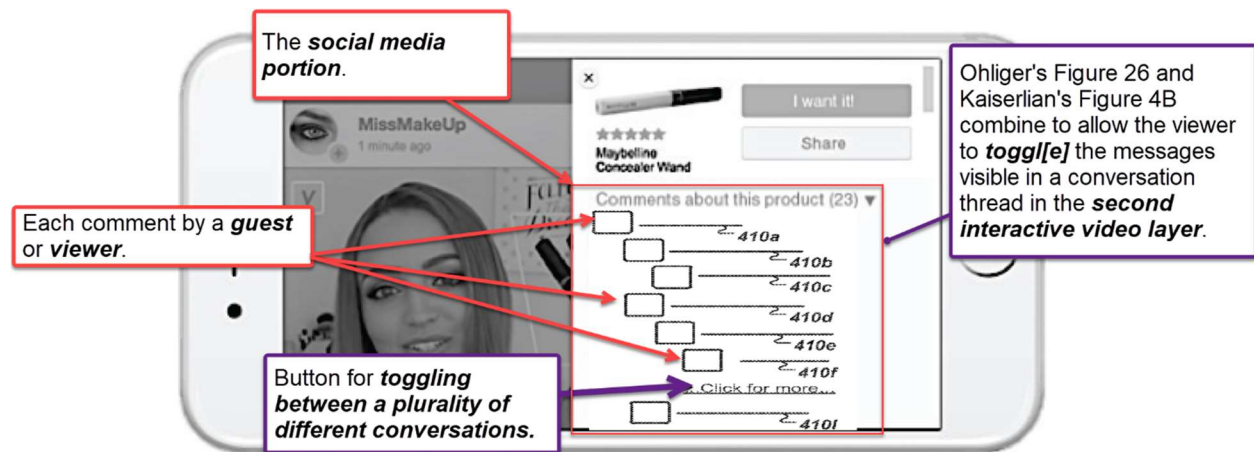
IPR of U.S. Patent No. 11,134,316

Decl. of Dr. Kevin C. Almeroth

455, and the indicator 470 permits a user to expand the portion 455 so that the messages 410g-410k are visible.” EX1014, 15:38-41.

298. The combined system would allow for reply comments within each video object’s conversation to be collapsed so that they are not visible unless the viewer interacts with the indicator to expand the reply comments. The resulting system renders all challenged claims obvious.

299. For example, a mockup combining Ohliger Figure 26 and Kaiserlian Figure 4B discloses the same functionality accused in Patent Owner’s Infringement Contentions.



### 3. Motivation to Combine & Reasonable Expectation of Success

300. In my opinion, a POSITA would have been motivated to apply Kaiserlian’s teachings to Ohliger’s video overlay system, and would have had a reasonable expectation of success in combining the references.

301. In my opinion, as a threshold matter, Ohliger, Kaiserlian, and the ’316 Patent are all analogous art because the ’316 Patent and the prior art references are

all incorporate electronic communications between individuals in written conversations. For example, the '316 Patent enables “the user to join a conversation with anyone watching the video media”. EX1001, 20:33-39. Ohliger and Kaiserlian similarly relate include messaging systems for users to communicate on written conversations.

302. Ohliger discloses both commenting and direct messaging between users functionalities. EX1006, 11:11-19, Figure 15. Kaiserlian directed to “presenting related communications” in a conversation thread. EX1014, 1:14-30.

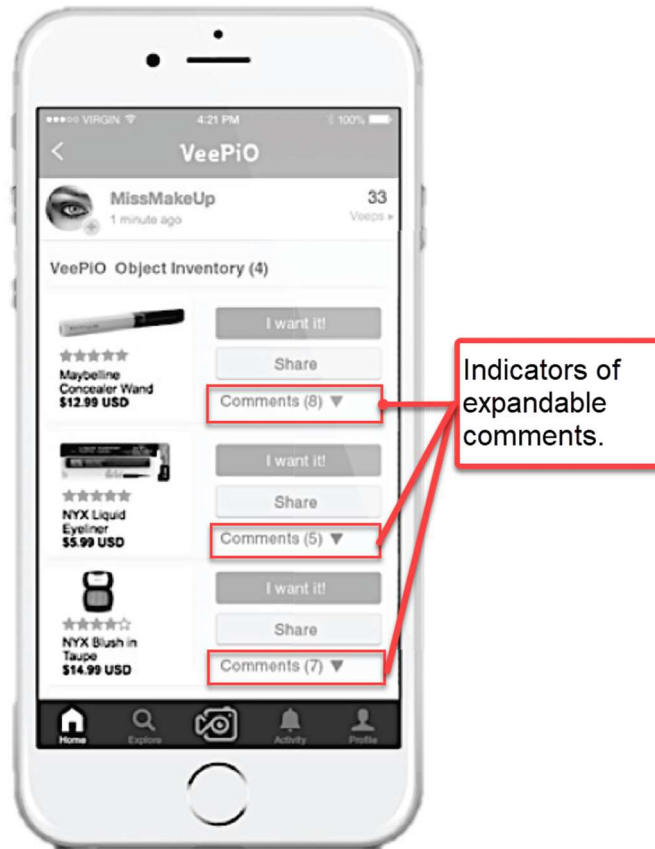
303. In my opinion, a POSITA would have looked to Ohliger and Kaiserlian when considering how to implement known systems for integrating viewer communications into interactive video content. In my opinion, a POSITA would have further been motivated to combine Ohliger and Kaiserlian for the following reasons.

304. ***Improved viewer experience:*** Within Ohliger, it was understood that Ohliger’s goal was to present user interfaces to provide compelling experiences for users to view and interact with video content. The focus on user engagement drives interface design decisions throughout the system. With that in mind, in my opinion, a POSITA would have been motivated to incorporate known ways of further improving user interfaces overlaid on video content, especially for smaller form-

factor devices with limited screen real estate, such as smartphones contemplated in Ohliger.

305. A POSITA would have recognized the opportunity to improve comment layouts through collapsible and expandable controls for portions of comment sets. This interface pattern addresses the fundamental challenge of displaying numerous comments on limited screen dimensions. The implementation enables viewers to manage visual complexity by selectively expanding conversations of interest while maintaining awareness of other available discussions.

306. For example, Ohliger includes such comment indicators on Figure 27's "Object Inventory" as highlighted below. The numerical indicators adjacent to each object represent collapsed comment sets that can be expanded on demand. A POSITA would have understood that this compact representation allows multiple conversation threads to coexist without overwhelming the display. The collapsible structure particularly benefits mobile viewing scenarios where screen space must accommodate both video content and interactive elements. Through selective expansion and collapse of comment sections, viewers maintain control over their visual experience while retaining access to all social interactions. Therefore, incorporating expandable comment controls represents a logical enhancement that optimizes Ohliger's interface for the mobile devices explicitly contemplated in the disclosure.



307. In my opinion, Kaiserlian discloses collapsible comment reply threading, a widely known interface design technique, that would have provided advantages in Ohliger's context that a POSITA would have been well familiar with. The hierarchical organization of comments addresses scalability challenges inherent in social discussions.

308. Instead of scrolling through lists of up to hundreds or thousands of comments, viewers would have been able to navigate between reply threads on comments of particular interest to viewers. By limiting the display of replies beyond a default depth, the user interface initially displays fewer comments. Further replies

are displayed only when the user selects a particular thread of replies for expansion.

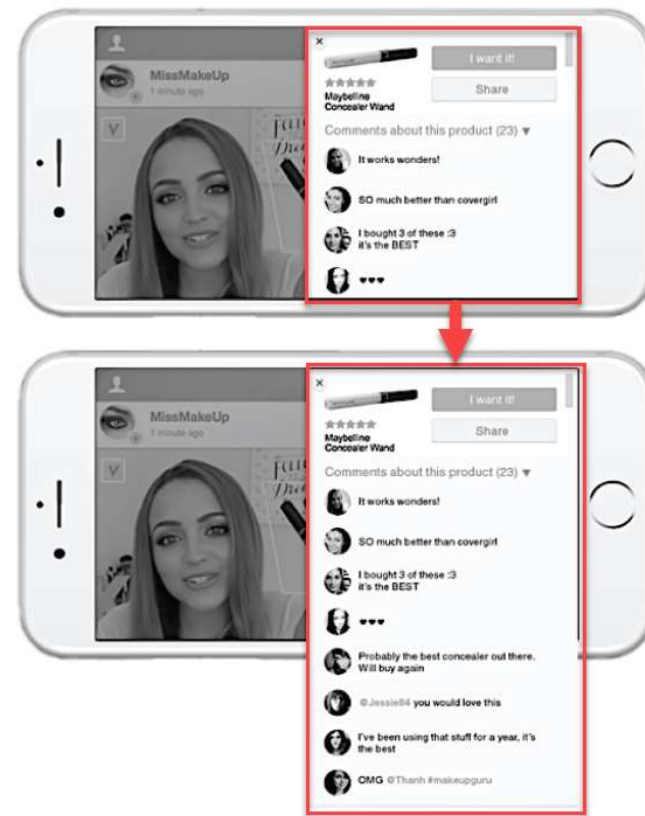
This mechanism preserves screen space while maintaining access to full discussions.

309. This enables the user to navigate the comments without needing to view all of the comments individually. A POSITA would have recognized that this organization is advantageous where the replies to a top-level comment often substantively relate to that top-level comment. The topical organization improves information discovery efficiency.

310. For items in videos, a POSITA would have recognized that this organization is advantageous because a viewer may not have interest in all of the topics discussed in the comments for a particular item. For example, if the item is an article of clothing, as often is the case in the examples of Ohliger, each viewer may have different topics for comments and questions for other viewers (colors, sizing, materials, cleaning, durability, etc.). A POSITA would have recognized that the diverse discussion threads benefit from structured organization.

311. A POSITA would have recognized that Kaiserlian's reply-based organization of these topics would allow for a more efficient and effective use of the comments for the viewers of the video. The implementation transforms linear comment lists into navigable topic hierarchies. Ohliger recognized the difficulty of large comment sets as seen in Figure 26 where Ohliger includes two images of the same screen with comments expanded beyond smartphone bounds:

Fig. 26  
Share & Comment



312. Kaiserlian provides improved viewer experiences by organizing comments into collapsible threads that hide comment threads unless viewers choose to expand threads. Combined with Ohliger, this would have allowed viewers to easily navigate through many comments even with limited screen space such as on smartphones. The space-saving benefit directly addresses Ohliger's documented interface challenges.

313. With Kaiserlian's superior organization and navigation over large comment lists, viewers would have more easily contributed to communities which

Ohliger recognizes as desirable. EX1006, 28:42-45. A POSITA would have understood that reducing navigation friction increases user participation rates. The hierarchical threading creates logical entry points for new contributors to join specific sub-discussions rather than attempting to parse a large set of undifferentiated comments.

314. Additionally, the expandable nature of those replies would have still allowed for the full set of comments to be accessible to each viewer, balancing initial simplicity with comprehensive access. Viewers control their level of engagement—scanning top-level comments for topics of interest, then selectively drilling into relevant discussion threads.

315. The technical implementation involves minimal modification to Ohliger's existing comment system. Adding collapse/expand indicators and maintaining thread state requires only incremental changes to the overlay rendering logic. Therefore, incorporating Kaiserlian's collapsible threading represents a predictable enhancement that solves Ohliger's acknowledged display constraints while preserving full interactive capabilities. The combination delivers superior navigation, improved screen utilization, and enhanced user engagement—all critical factors for the mobile viewing scenarios central to Ohliger's system.

316. *Ohliger already contemplates separating comments by topic*: Ohliger describes “sub-comments” as comments particularly for “a specific interactive

object in the video ... as opposed to comments that relate to the video as a whole.”

EX1006, 11:14-19. As Ohliger’s Figure 27 shows, each video object has a distinct set of comments that a viewer may observe independent of any other video object’s comments. The system already implements first-level topical separation through object association. A POSITA would have recognized the analogous improvement of an integrated, single comment thread organized with the collapsible sub-threads as in Kaiserlian. The enhancement extends Ohliger’s existing topical structure to a second hierarchical level. This two-tier organization creates object-level topics with conversation-level subtopics. A POSITA would have understood that viewers could first select an object of interest, then navigate within that object’s discussion through collapsible reply threads addressing different topics—sizing questions branching into one thread, material queries into another, styling suggestions into a third.

317. *Software modification would be predictable*: In my opinion, changes to the comments overlay of Ohliger would have been predictable and minimally impactful on the overall system given the modularity of the design and the ability to implement the changes in a specific component of the user interface design. The isolated nature of UI modifications limits cascading effects on other system components.

318. A POSITA would have recognized that comment display logic operates independently from core video processing and object identification functions. The

modification involves adjusting how existing comment data is rendered rather than altering underlying data structures or server communications. Implementing collapsible threading requires adding state management for expanded/collapsed status and modifying interface to display hierarchical relationships.

319. The user interface modifications are limited to modifications to the viewer software and thus predictable due to their limited scope within Ohliger's system. A POSITA would have understood that existing user interface frameworks commonly used for mobile applications would have readily supported the dynamic show/hide functionality and hierarchical display of comments.

#### **4. Analysis**

320. The analysis for Ground 3 incorporates by reference the analysis for Ground 1. Below, Petitioner provides additional analysis for individual claim limitations that further rendered obvious by the proposed combination of Ohliger and Kaiserlian (in addition to being rendered obvious by Ohliger alone).

##### **(a) Claim 1**

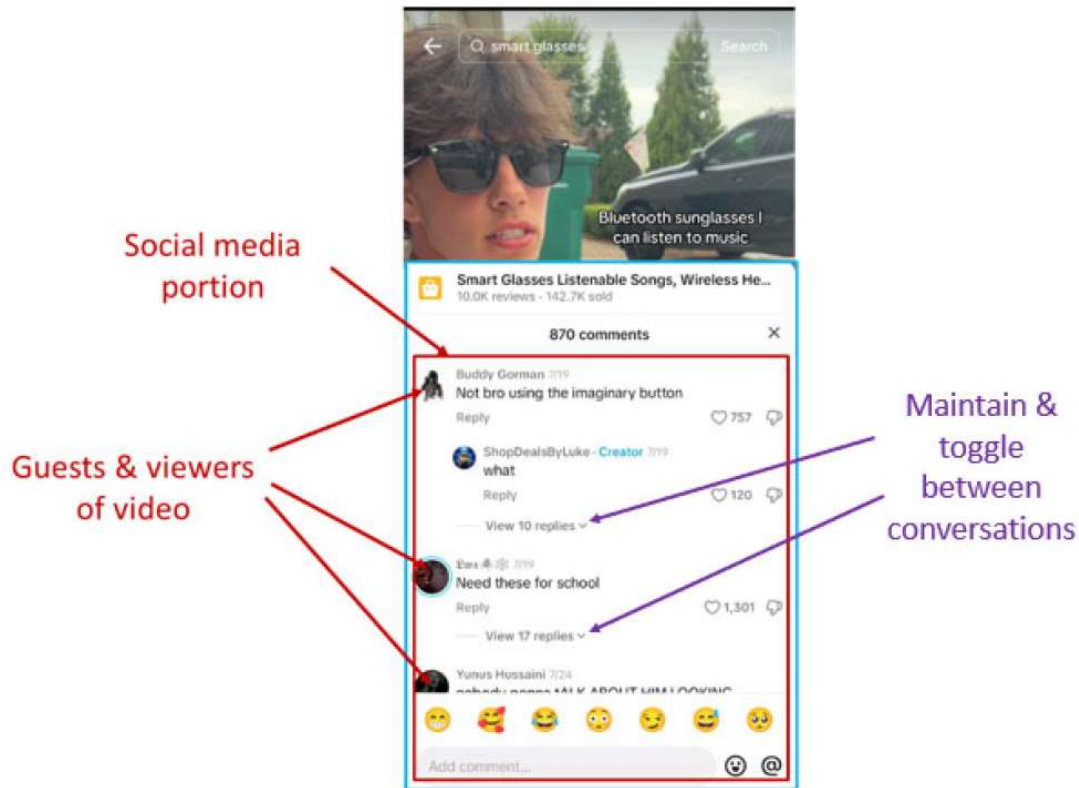
- (i) 1[c][vi]: wherein the user is capable of maintaining and toggling between a plurality of different conversations displayable in the social media portion in the second interactive video layer as an overlay to the video media displayed by the first media playback device,**

321. In my opinion, the proposed combination of Ohliger and Kaiserlian discloses this limitation. Ohliger discloses comments for each video object

appearing in a video (*conversations displayable in the social media portion in the second interactive video layer as an overlay to the video media displayed by the first media playback device*). See section VII(A)(2)(x) – (xiii); EX1006, Fig. 26.

These object-specific comment threads establish the foundation for multi-conversation display.

322. Further, Patent Owner's Infringement Contentions in the District Court litigation for claim 1's *maintaining and toggling between a plurality of different conversations* accuse the expansion of a conversation thread's replies as seen in the figure below. EX1005, 48. According to Patent Owner's interpretation, expanding and collapsing conversation threads constitutes toggling between conversations. In my opinion, the proposed modification of Ohliger with Kaiserlian discloses the same functionality accused in Patent Owner's Infringement Contentions.

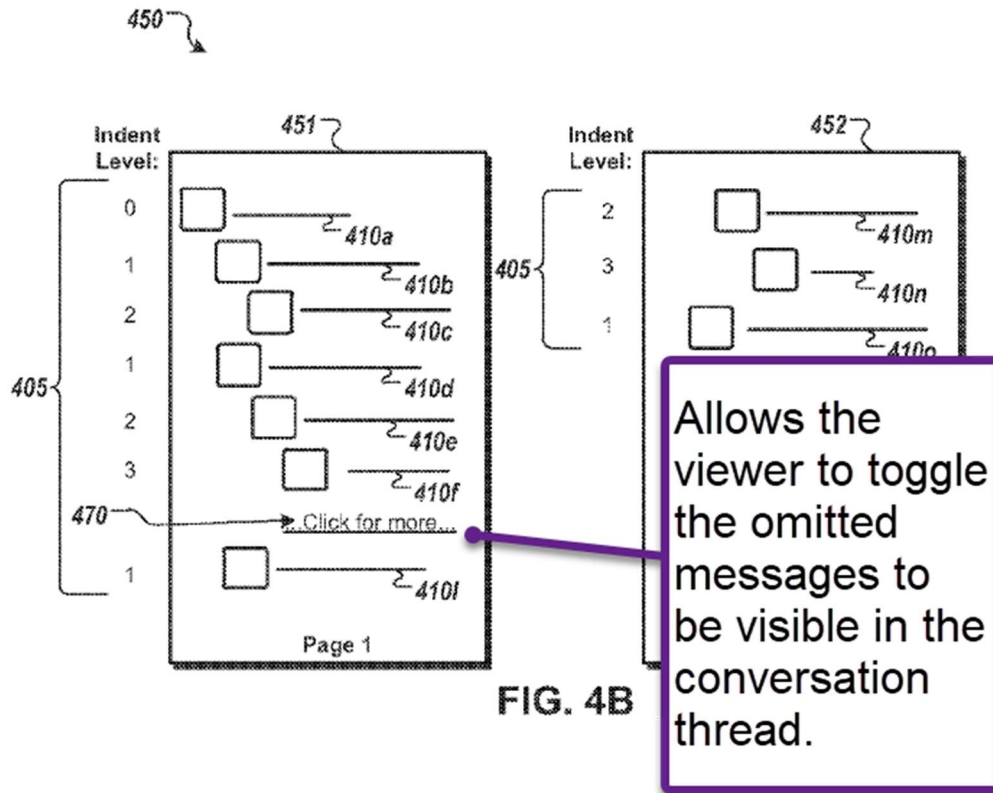


323. Kaiserlian's collapsible threading, when applied to Ohliger's object-specific comments, creates multiple expandable conversation branches within each object's discussion. The hierarchical structure organizes replies into navigable subtopics. A POSITA would have recognized that viewers can toggle between these conversation threads under Patent Owner's interpretation by selectively expanding replies of interest while keeping others collapsed. The implementation enables maintaining awareness of multiple conversation topics through visible thread headers while controlling which full discussions are displayed. This selective display mechanism optimizes screen utilization while preserving access to all discussions.

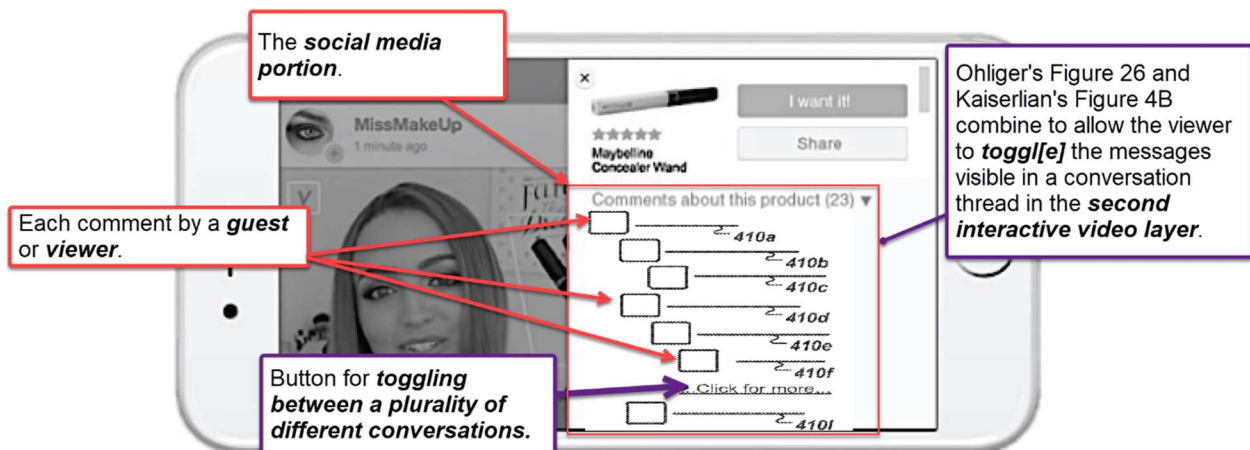
324. In my opinion, Kaiserlian discloses the same expansion of a conversation thread's replies that Patent Owner accuses of infringement. First, Kaiserlian discloses "omit[ting]" or hiding "messages in a conversation thread ... based on ... indent levels" where the indent level is indicative a reply thread. EX1014, 1:25-41. The indentation visually represents the hierarchical relationship between comments and their replies.

325. Kaiserlian's Figure 4B includes a "[c]lick for more..." button that "permits a user to expand ... the messages" that were previously omitted. EX1014, 15:38-41; 15:31-38 ("an ellipsis or other symbol[] ... indicates that the portion 455 is omitted" and "a user can interact with [it] to cause the omitted portion 455 to be presented."). This interactive control provides the toggling mechanism between collapsed and expanded states.

326. Therefore, under Patent Owner's interpretation, the viewer can use the click for more button for ***maintaining and toggling between a plurality of different conversations***. The combination of Ohliger's object-specific comment threads with Kaiserlian's expandable reply threading creates the exact functionality Patent Owner identifies as meeting this limitation. Through selective expansion and collapse of conversation branches, viewers control their engagement with multiple discussion threads while maintaining the overlay presentation on the video media.



327. In my opinion, when those teachings are applied to Ohliger's sub-comments for specific video objects, the combination renders obvious the claimed *toggle*, as shown in the following mock-up:



**(b) Claims 17 & 18**

328. In my opinion the proposed combination of Ohliger and Kaiserlian further renders limitations 17[b][vi] and 18[d][vi] obvious for the reasons set forth above for limitation 1[c][vi].

**D. Ground 4: Ohliger and Tapissier**

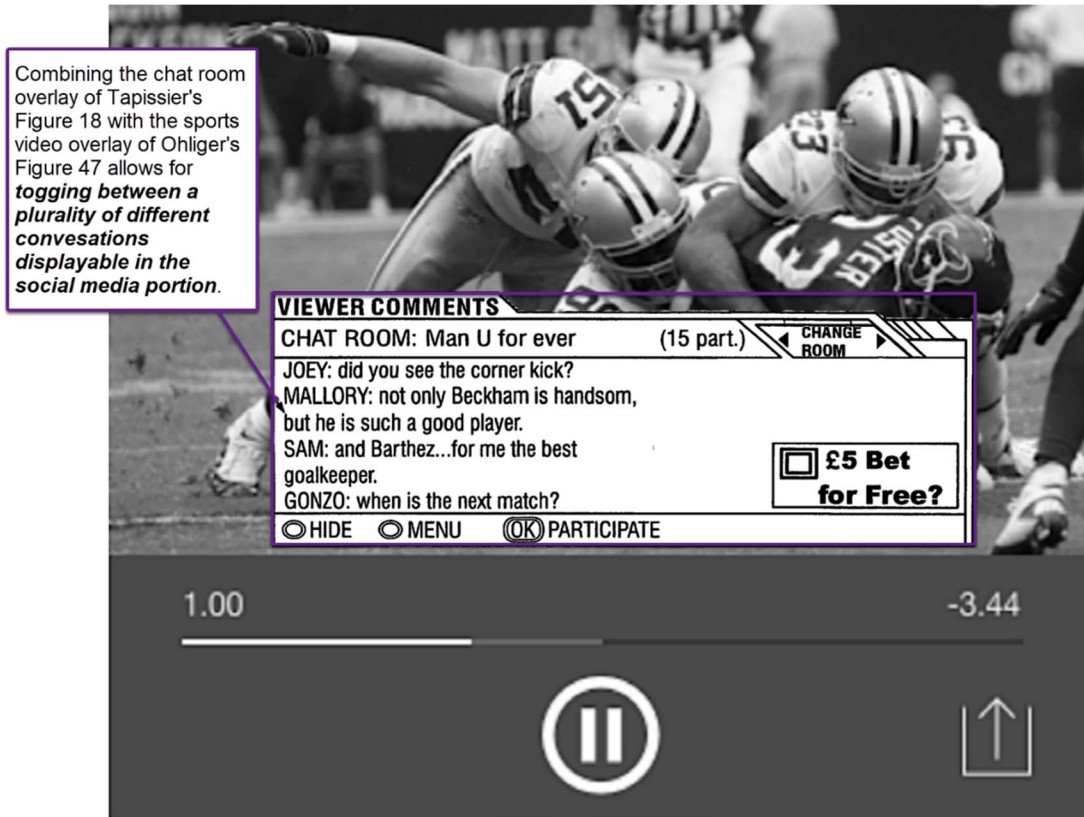
**1. Background**

329. E.P. No. 1,425,918 (“Tapissier”) was filed September 12, 2002 and published June 9, 2004. Tapissier is prior art under 35 U.S.C. 102(a). Tapissier describes “a method for broadcasting a chat room in an interactive television environment.” EX1007, [0007].

**2. Proposed Combination**

330. Ohliger’s system would be modified, according to the teachings of Tapissier, to combine Tapissier’s overlaid “[c]hat rooms” where “the Chat user has the option of selecting participation in a chat room from a list of Chat rooms” with Ohliger’s interactive videos. EX1007, [0064]. As part of this modification, Ohliger’s interactive videos would include the chatroom overlay, as in Tapissier’s Figure 18, to allow for chat room functionality with toggling between different chatrooms associated with a particular video. EX1007, Fig. 18 The combined system would expand the existing interactive overlays of Ohliger to include the chat room overlay of Tapissier with it enabling a viewer to toggle between different chatrooms. The resulting system renders all challenged claims obvious. For example, consider the

mockup combinations below based on Ohliger's Figures 47, and Tapissier's Figure 18.



### 3. Motivation to Combine & Reasonable Expectation of Success

331. In my opinion, a POSITA would have been motivated to apply Tapissier's teachings to Ohliger's video overlay system, and would have had a reasonable expectation of success in combining the references.

332. As a threshold matter, Ohliger, Tapissier, and the '316 Patent are all analogous art because the '316 Patent and the prior art references are all incorporate electronic communications between individuals in written conversations as overlays on video content. For example, the '316 Patent enables "the user to join a

conversation with anyone watching the video media”. EX1001, 20:33-39. Ohliger and Tapissier similarly relate include messaging systems for users to communicate in written conversations that are overlaid on video content.

333. Ohliger discloses both commenting and direct messaging between users functionalities. EX1006, 11:11-19; Fig. 15. Tapissier relates to “communications in the interactive television environment ... providing a chat room.” EX1007, [0001], Fig. 18.

334. In my opinion, a POSITA would have looked to Ohliger and Tapissier when considering how to implement known systems for integrating viewer communications into interactive video content. In my opinion, a POSITA would have further been motivated to combine Ohliger and Tapissier for the following reasons.

335. *Improved viewer experience*: Within Ohliger, social interactions amongst the community of viewers is recognized as desirable. EX1006, 28:42-45. With that in mind, in my opinion, a POSITA would have been motivated to incorporate known ways of further improving the social interactions through the overlay of user interfaces on the video content.

336. In my opinion, a POSITA would have known of extensive social interactions enabled by Tapissier’s overlay of chat rooms on video content. Tapissier’s multi-room architecture provides simultaneous parallel conversations.

Integrating Tapissier's multiple chat room overlay into the interactive videos of Ohliger would have created an improved social media experience for viewers. The combination would have enabled richer social dynamics beyond single-threaded discussions.

337. The viewer would have been enabled to engage with the community of viewers in a live, robust set of chatrooms in real-time while video content is simultaneously viewed. EX1007, [0051]. This real-time interaction transforms passive viewing into active social participation. A POSITA would have understood that synchronous communication creates immediate engagement opportunities that more asynchronous comments cannot match.

338. A POSITA would have recognized that for particular types of video content, like the sports video in Ohliger's Figure 47, a category of content for which live viewing was common, it would have been desirable to integrate the same real-time conversations that Tapissier disclosed for sporting events in Figure 18. A POSITA would have known that temporal alignment between live events and viewer reactions enhances the viewing experience.

339. Therefore, incorporating Tapissier's chat room overlays into Ohliger's interactive video system addresses the recognized desire for enhanced social interaction while maintaining visual access to the video content. The combination

leverages proven real-time communication mechanisms to create a more engaging social viewing environment aligned with Ohliger's community-focused objectives.

340. ***Commercial benefit from increased interactivity:*** A POSITA would have an economic motivation to integrate the chat room features from Tapissier into the interactive viewer in Ohliger given that the opportunity to interact with other viewers would increase the time the user spends in the interactive viewer, which will subsequently increase the possibility the user purchases a product from a retailer, brand, affiliate, or advertiser. EX1006, 11:32-37 (“An object of the present invention is to take conventional video content analysis technology ... and apply it to the field of marketing communication, advertising and commerce transaction.”), 10:53-60. Ohliger discloses advertising content such as links to e-commerce websites where the products in the video may be purchased. EX1006, 12:29-37. A POSITA would have recognized that advertisers would have placed a higher value on advertisements that viewers viewed and engaged with for an extended period of time. Further, Ohliger distinguishes its advertising as including both “push” advertising where ads are sent to the viewer and “pull” advertising where ads are requested by the viewer through interaction with video content such as the “I want this” button. EX1006, 14:31-59.

341. A POSITA would have recognized that the addition of chat rooms would have increased the time the viewer spent viewing Ohliger's push

advertisements and increased the interactions by the viewers that would lead to more pull advertisements. For example, the social engagement maintains viewer attention with the video. Therefore, integrating Tapissier's chat room functionality serves Ohliger's commercial objectives by extending viewing sessions, increasing ad exposure, and fostering social dynamics that promote product interest and eventual purchase. The economic benefits align with Ohliger's stated goal of applying video analysis technology to commerce transactions.

342. *User data collection*: In my opinion, a POSITA would have had an economic motivation to integrate the chat room feature from Tapissier into Ohliger. These viewer chats—being user data—would have been valuable to advertisers as it allows for targeted advertising, which increases the probability a user clicks on an advertisement. EX1007, [0026]. The chat data provides rich behavioral insights beyond basic viewing patterns.

343. An integral part of Ohliger is to auction rights to hyperlinks to third parties. EX1006, 1:27-33. The auction model depends on maximizing the value of each hyperlink placement. Thus, in my opinion, a POSITA would have understood that the ability to provide additional data to advertisers—facilitating targeted advertising—would likely increase the profit made from each auctioned hyperlink. EX1006, 14:12-21.

344. The chat room conversations reveal real-time viewer interests, preferences, and purchase intent. A POSITA would have recognized that analyzing chat content enables sophisticated audience segmentation. Advertisers gain insights into viewer demographics, product opinions, and social influences that shape purchasing decisions. This granular data justifies higher auction bids as advertisers can more accurately predict conversion rates.

345. The technical implementation would have involved capturing and processing chat data alongside existing viewer profile information. Ohliger already tracks viewer interactions and builds profiles for targeted advertising. Adding chat room data enriches these profiles with conversational context and peer interactions. Therefore, integrating Tapissier's chat functionality provides economic value through enhanced data collection that increases auction revenues—a direct alignment with Ohliger's monetization strategy based on auctioning hyperlinks to the highest bidder.

346. *Same field of art:* Ohliger and Tapissier relate to electronic communications as a part of social media interactions overlaid on interactive videos. EX1006, 3:9-20; EX1007, [0001]. Further, both consider interactive television. EX1006, 15:41-44, 24:1-6; EX1007, [0002], [0005]. Both contemplate messages or comments between a group of individuals. Both contemplate using networking technologies as parts of the system. EX1006, 3:21-23; EX1007, [0058].

347. *Ohliger already contemplates messaging between viewers and overlaying viewer comments on video content:* Ohliger discloses direct messaging between viewers and overlaying viewer comments on video content. EX1006, Fig. 15, Fig. 26. These existing communication mechanisms establish the technical foundation for enhanced social features. Ohliger already discloses “sub-comments” related to particular objects in a video. EX1006, 11:14-19, Fig. 27. The sub-comment demonstrates the system’s capability to manage multiple distinct conversation threads.

348. Through sub-comments, Ohliger already contemplates overlaying multiple conversations amongst viewers for a particular video. Each object maintains its separate discussion thread, proving the system can handle parallel conversation streams. A POSITA would have recognized that the infrastructure supporting object-specific sub-comments readily extends to support topic-based chat rooms.

349. In my opinion, a POSITA would have reasonably expected to succeed in adding overlaid chat room messaging as in Tapissier due to its commonality with existing features of Ohliger. The technical similarities reduce implementation complexity. Both systems overlay text-based communications on video content. Both manage multiple conversation contexts. Both enable real-time viewer interaction.

350. The modification would have involved adapting Ohliger's existing overlay rendering and message handling to accommodate Tapissier's chat room structure. Ohliger's server already processes and distributes viewer comments; adding chat room channels represents an incremental expansion of this capability. The client-side interface already displays overlaid conversations; presenting multiple chat rooms would have required only modest UI adjustments. Therefore, the substantial overlap between Ohliger's implemented features and Tapissier's chat room functionality ensures a reasonable expectation of success, as the combination leverages proven components rather than requiring novel technical solutions.

351. *User interface modification would be predictable*: Modifications to Ohliger's overlaid user interfaces for overlaid chat rooms of Tapissier would have been predictable and minimally impactful on the overall system given the modularity of the design, existing messaging and overlay functionalities, and the ability to implement the changes in a specific component of the user interface design. The architectural similarities facilitate straightforward integration.

352. A POSITA would have recognized that Ohliger's system already contains the core components necessary for chat room implementation: real-time messaging infrastructure, overlay rendering capabilities, and multi-conversation management through sub-comments. The addition of chat rooms leverages these existing subsystems rather than requiring fundamental architectural changes.

353. Further, the user interface and chat room modifications are limited to modifications to the viewer software and thus predictable due to their limited scope within Ohliger's system. The technical implementation would have followed established patterns. Message transport mechanisms already handle direct messages and sub-comments; adapting them for chat room broadcasts involves routine protocol extensions. The overlay system already positions multiple UI elements over video content; adding chat room panels uses the same positioning and rendering logic. State management for multiple conversations exists for sub-comments; extending it to chat rooms requires only incremental data structure modifications.

354. Therefore, the combination of Ohliger and Tapissier represents a predictable variation achievable through localized modifications to existing components. The modularity of Ohliger's design, combined with the functional overlap between its current features and Tapissier's chat rooms, ensures minimal system disruption while delivering enhanced social interaction capabilities.

#### **4. Analysis**

355. The analysis for Ground 4 incorporates by reference the analysis for Ground 1. Below, Petitioner provides additional analysis for individual claim limitations that further rendered obvious by the proposed combination of Ohliger and Tapissier (in addition to being rendered obvious by Ohliger alone).

**(a) Claim 1**

- (i) 1[c][vi]: wherein the user is capable of maintaining and toggling between a plurality of different conversations displayable in the social media portion in the second interactive video layer as an overlay to the video media displayed by the first media playback device**

356. In my opinion, the proposed combination of Ohliger and Tapissier discloses this limitation. Ohliger discloses comments for each video object appearing in a video. *See* section VII(A)(2)(x) – (xiii); EX1006, Fig. 26. These object-specific comment threads provide the base conversation structure.

357. Tapissier describes a chat room overlay on video broadcasts such as for interactive television. EX1007, [0001]. Tapissier's chat rooms enables multiple simultaneous conversation streams. Tapissier further discloses that "[t]he Chat user has the option of selecting participation in a chat room from a list of Chat rooms." EX1007, [0064], [0011]. This selection mechanism provides user control over conversation access and participation.

358. Tapissier's Figure 18 includes a button for ***toggling*** between chat rooms. EX1007, Fig. 18. A POSITA would have recognized that this toggling functionality directly enables switching between different active conversations. The interface allows viewers to maintain awareness of multiple chat rooms while actively participating in selected discussions.

359. When combined with Ohliger's system, Tapissier's multi-room chat overlay creates parallel conversation channels overlaid on the video content. Viewers can toggle between various chat room discussions. Therefore, the combination enables *maintaining and toggling between a plurality of different conversations* through both Ohliger's object-specific sub-comments and Tapissier's selectable chat rooms. The overlay presentation of these multiple conversation streams on the video media satisfies the requirement for conversations *displayable in the social media portion in the second interactive video layer as an overlay to the video media*.

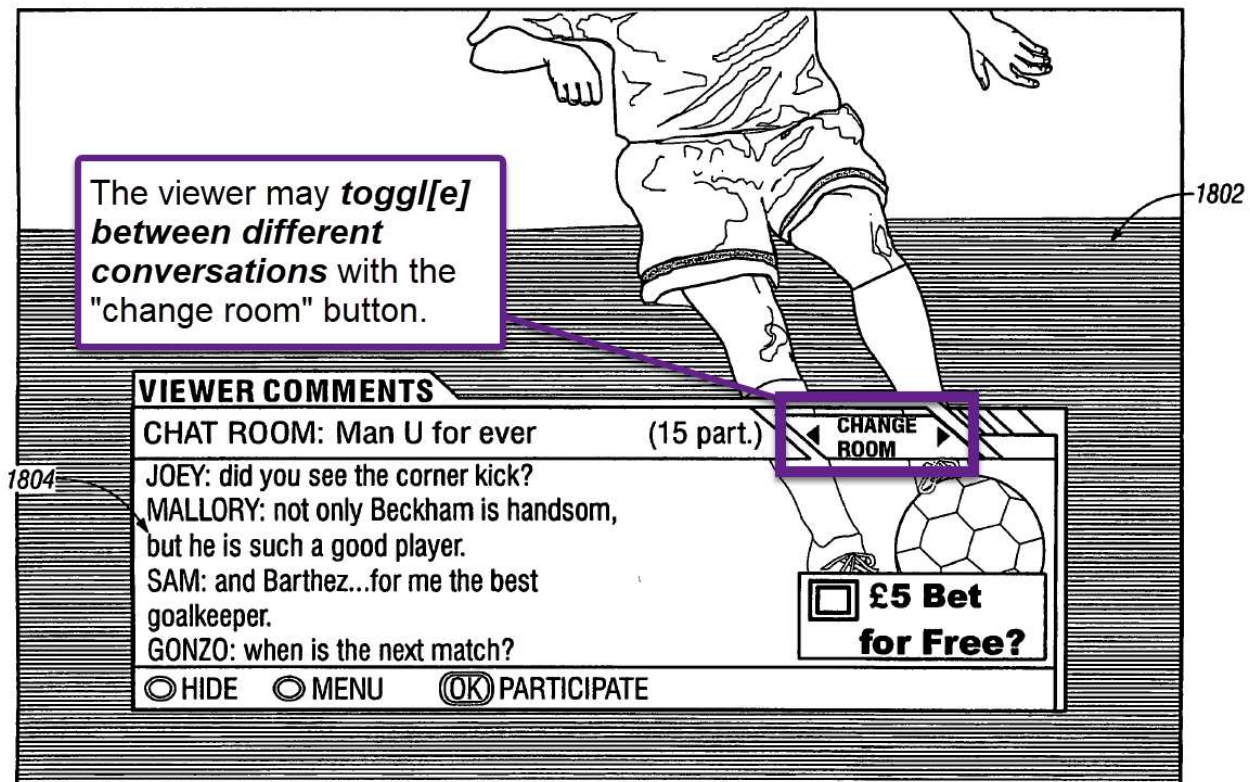


FIG. 18

**(b) Claim 2**

360. In my opinion, Tapissier discloses that its chat room overlay can be turned on or off. EX1007, [0064] (“The Chat user has the option of selecting participation in a chat room from a list of Chat rooms or selecting a chat room associated with a given broadcast.”), [0066] (“When presented on a video channel, a pop-up menu invites the viewer to join the chat sessions ... [i]f the client wants the pop-up to appear again ... this is selectable by the client”), Fig. 17 (“Exit” button), Fig. 18 (“Hide” button). A POSITA would have recognized that Tapissier’s “Exit” button terminates chat participation while the “Hide” button conceals the overlay without disconnecting from the conversation. The pop-up invitation system allows users to decline initial participation or rejoin after exiting. This granular control addresses different viewing preferences—some users may want continuous social interaction, others may prefer intermittent participation, and some may choose to avoid chat entirely. Therefore, through Tapissier’s exit, hide, and participation selection features integrated with Ohliger’s overlay system, the combination enables users to *turn on or off* the social media portion of the *second interactive video layer* based on their viewing preferences.

**(c) Claim 16**

- (i) 16[a][ii]: wherein the plurality of different conversations comprises at least a public conversation which enables all viewers of the video media to join and a private conversation**

**which enables a permitted subset of all viewers of the video media to join.**

361. In my opinion, the proposed combination of Ohliger and Tapissier discloses this limitation. Ohliger discloses overlaying comments for each video object appearing in a video (*a public conversation which enables all viewers of the video media to join*). See section VII(A)(2)(x) – (xiii); EX1006, Fig. 26. The open access nature of sub-comments allows any viewer to participate in object-specific discussions.

362. Tapissier describes a chat room overlay on video broadcasts including both public chat rooms and “private room[s]” where “users ... may exchange private messages” (*a private conversation which enables a permitted subset of all viewers of the media to join*). EX1007, [0064]; [0001]. The dual-mode conversation architecture supports both open and restricted communication channels.

363. A POSITA would have recognized that combining these systems creates comprehensive social interaction capabilities. The public chat rooms and Ohliger’s sub-comments provide community-wide discussion forums accessible to *all viewers*. Meanwhile, Tapissier’s private rooms establish exclusive communication channels limited to a *permitted subset of all viewers*. Therefore, through Ohliger’s public sub-comments combined with Tapissier’s dual public/private chat room structure, the combination discloses both *a public conversation which enables all viewers of the video media to join* and *a private*

*conversation which enables a permitted subset of all viewers of the media to join,*

satisfying both aspects of this limitation.

**(d) Claim 17**

364. Tapissier further renders limitation 17[b][vi] obvious for the reasons set forth above for limitation 1[c][vi].

**(e) Claims 18**

- (i) 18[d][i]: d) a software module for the generation of a second interactive layer to display with the video media and optionally the first interactive layer,**

365. In my opinion the proposed combination of Ohliger and Tapissier discloses this limitation. Tapissier describes a chat room overlay (*second interactive layer to display with the video media*) on video broadcasts. EX1007, [0001].

366. In the proposed combination with Ohliger, Ohliger's "V" button would remain, thereby allowing the user to *optionally* display Ohliger's animations, outlines, and halos shown as clickable overlays on top of video objects (*first interactive layer*) alongside the underlying video content (*video media*) and Tapissier's chat room overlay (*second interactive layer*). The preserved toggle functionality maintains user control over layer visibility.

367. A POSITA would have recognized that the combination supports independent control of each layer. Users could display: (1) video only, (2) video with object overlays, (3) video with chat overlay, or (4) video with both overlays

simultaneously. The “V” button controls the *first interactive layer* while Tapissier’s hide/exit controls manage the *second interactive layer*. This granular control ensures viewers customize their experience based on immediate preferences. Therefore, through Ohliger’s toggle-controlled object overlays combined with Tapissier’s selectable chat rooms, the combination enables *optionally* displaying the *first interactive layer* alongside the *video media* and *second interactive layer*. See also Ground 1, limitation 18[d][i] for discussion of other aspects of this claim limitation.

**(ii) 18[d][vi]**

368. Tapissier further renders limitation 18[d][vi] obvious for the reasons set forth above for limitation 1[c][vi].

**(f) Claims 19**

369. Tapissier further renders claim 19 obvious for the reasons set forth above for claim 16.

**E. Ground 5: Ohliger, Sakhartov, and Kaiserlian or Tapissier**

370. For the reasons set forth in Grounds 3-4, in my opinion, it would further be obvious to combine Ohliger and Sakhartov (Ground 2) with either Kaiserlian (Ground 3) or Tapissier (Ground 4) to arrive at the challenged claims of the ’316 patent.

## **VI. Secondary Considerations of Non-Obviousness**

371. I understand that so-called “secondary considerations” are legally relevant to an obviousness analysis. It is my opinion that the secondary considerations I consider here further weigh in favor of obviousness.

372. Any secondary considerations further weigh in favor of obviousness, and certainly do not overcome the strong showing of obviousness. First, there was no long felt, unresolved need met by the '316 patent: to the extent wide-scale deployment of interactive video systems was delayed in the United States, it was because of industry and market barriers, not technical barriers. Further, the interactive video systems and other art discussed in the State of the Art section, which predate the '316 patent's priority date, indicate that many within the market observed the desirability of such systems and were concurrently working towards implementing them.

## **VII. Conclusion**

373. In signing this declaration, I recognize that the declaration will be filed as evidence in a contested case before the Patent Trial and Appeal Board of the United States Patent and Trademark Office. I also recognize that I may be subject to cross-examination in the case and that cross-examination will take place within the United States. If cross-examination is required of me, I will appear for cross-examination within the United States during the time allotted for cross-examination.

\* \* \*

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on the information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code.

Respectfully submitted,

Date: September 2, 2025



Kevin C. Almeroth, Ph.D.